Characterizing the human vaginal microbiome using high-throughput sequencing

Jean Megan E. Macklaim, *The University of Western Ontario*

Supervisor: Greg Gloor, *The University of Western Ontario*
Joint Supervisor: Gregor Reid, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biochemistry
© Jean Megan E. Macklaim 2013
CHARACTERIZING THE HUMAN VAGINAL MICROBIOME USING HIGH-THROUGHPUT SEQUENCING

(Thesis format: Integrated Article)

by

Jean Megan Elizabeth Macklaim

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Jean M. E. Macklaim 2013
Abstract

The human vaginal microbiome undoubtedly has a significant role in reproductive health and for protection from infectious organisms. Recent efforts to characterize the bacterial species of the vagina using molecular techniques have uncovered an unexpected diversity. Using high-throughput sequencing I sought to describe the structure and function of the vaginal microbiome under different physiological states including healthy, bacterial vaginosis (BV), post-menopausal vaginal atrophy, and acute vulvovaginal candidiasis (VVC).

Partial 16S rRNA gene sequencing revealed that healthy, asymptomatic women most often have vaginal biotas dominated by *Lactobacillus iners* or *L. crispatus*. In contrast, BV is a heterogeneous, highly diversified condition with reduced *Lactobacillus* abundance. Similar to BV, post-menopausal women experiencing vaginal dryness were depleted in lactobacilli and had a more diverse vaginal profile. In the case of VVC, the biotas were not significantly altered compared to healthy women despite the fungal overgrowth.

One organism, *Lactobacillus iners* was ubiquitously present in all conditions, and became predominant following antibiotic and probiotic treatment of BV. To uncover the potential role of this bacterium, I used whole genome sequencing of vaginal isolate AB-1. The genome is predicted to be the smallest of any *Lactobacillus* at 1.3 Mbp, but having a higher proportion of horizontally acquired genes. These results, along with predicted adhesins and a cholesterol-dependent cytolysin, indicate *L. iners* is highly adapted for the vagina and could have an uncharacterized role in the etiology of BV.

As BV is the most common vaginal ailment with severe implications on acquisition and transmission of sexually transmitted infections, and complications during pregnancy, I sought
to examine the functional contribution of the organisms during BV using meta-RNA sequencing. *L. iners* drastically modulates gene expression in response to BV, and notably increases expression of a cholesterol-dependent cytolyisin, mucin and glycerol transport and metabolic enzymes, and genes belonging to a CRISPR system - suggestive of bacteriophage influence in the community. Although diverse in taxonomic membership, there is evidence of functional conservation in BV including preference for glycogen and glycerol as carbon sources, and predicted end products of metabolism including an abundance of succinate and short-chain fatty acids. These studies add significantly to our understanding of the role lactobacilli can play in vaginal and reproductive health.

**Keywords**

Human microbiome, bacterial vaginosis, women’s health, next-generation sequencing, 16S tag sequencing
Co-Authorship Statement

The experiments and data analyses presented in this thesis were primarily conducted by Jean Macklaim with Greg Gloor and Gregor Reid. The published manuscripts were written by Jean Macklaim, Greg Gloor, and Gregor Reid. Exceptions are noted below.

Chapter 1: Vaginal microbiome and epithelial gene expression of post-menopausal women with moderate to severe dryness

Ruben Hummelen and Gregor Reid conceived of the study design and collected the samples. Amy McMillan extracted sample DNA and PCR amplified. Jo-Anne Hammond connected clinical samples. Rebecca Vognsa and David Koenig contributed data analysis. Jordan Bisanz performed analysis of the human microarray data. Jean Macklaim performed analysis of the microbiota data.

Chapter 3: At the crossroads of vaginal health and disease: the genome sequence of Lactobacillus iners AB-1

Gregor Reid and Greg Gloor conceived of the study design. Kingsley Anukam prepared the samples for sequencing. Jean Macklaim and Greg Gloor performed the genome assembly, annotation, and analyses, and the protein extraction for mass spectroscopy. Jean Macklaim and Sarah Cribby performed the immunogold labeling and TEM images with Judy Sholdice.

Chapter 4: Comparative meta-RNA-seq of the vaginal microbiota and differential expression by Lactobacillus iners in health and dysbiosis


Chapter 5: Effect of antimicrobial and probiotic therapy on the vaginal microbiota

The initial clinical trials were conceived and performed by Martinez, Gregor Reid, and Elaine Cristina Pereira De Martinis. The DNA extractions and PCR amplifications were performed by Amy McMillan. Jean Macklaim performed the primary data analysis with Greg Gloor. Rob Knight and Jose Clemente aided in data analysis using QIIME. Jean Macklaim, Greg Gloor, and Gregor Reid wrote the manuscript.
Acknowledgments

I am very fortunate to have had two fantastic supervisors for my PhD thesis and I can't express my gratitude for their guidance on this journey. So thank you to Gregor Reid and Greg Gloor for taking the chance on me. I came in not knowing anything about microbiome, bacteria, or computational biology but your patient mentorship and guidance (and great sense of humour) made this thesis possible. You both gave me the freedom to explore and challenge my ideas - good and bad.

Greg: You've had seemingly unlimited patience, and you've gone above and beyond as a supervisor. You are not afraid to walk on the bleeding edge of science, and your innate curiosity about the world is the core of a good scientist.

Gregor: You have a fantastic ability to draw connections and inspiration from anywhere in your research. I especially want to thank you for pushing me beyond what I thought I was capable of. Your passion for making the world a better place is something everyone should have.

A few people have had significant influence on this thesis:

Ruben: you walked into our lab with over 300 samples from Africa and changed the course of our research for the years to come. Thank for your collaborative attitude, and for the many interesting discussions.

Andrew: I remember the day I was whining (yet again) about the woeful inadequacies of the bioinformatic tools for the data we were trying to analyze, and you took it as a challenge to make something better. One day I may eventually understand ALDEx. Thank you for you mentorship over the years, and for your healthy dose of cynicism.

Russ: I feel we've shared a camaraderie (or is it misery?) that can only be understood by fellow grad students. Your humour and scientific insight have been invaluable, and I thank you for your patient explanations on computers and programming.
Thank you to the Department of Biochemistry for providing a welcoming and collegial environment for research. To my committee members, John McCormick and David Edgell, I thank you for your always insightful input into my project. And to the Molecular Biology group for providing a friendly forum for scientific discussions.

To the past and present members of the Gloor lab where I spent most of my time: I thank you for your humour and friendship and the many intellectual discussions (once in a while even about science). "On second thought, let's not go to the Gloor lab - 'tis a silly place”.

To the Reid lab: So many of you have impacted this thesis in small but significant ways. I consider every one of you who have passed through in my time here a friend.

On a personal note, I thank my Mom, Dad, and sister Kristy for your love and support and for encouraging me pursue the things that fascinate me.

I also thank my karate Senseis for their influence (within and outside karate) and for showing the value of commitment.

Nikolaj: You have such a profound influence on my life and many of the steps I have taken to get here would not have happened had you not encouraged me. My gratefulness for your love and support cannot be put into words. I am thankful to be on this journey together with you.
# Table of Contents

Abstract ................................................................................................................................. ii  
Co-Authorship Statement ........................................................................................................ iv  
Acknowledgments .................................................................................................................. v  
Table of Contents .................................................................................................................. vii  
List of Tables ........................................................................................................................ xii  
List of Figures ......................................................................................................................... xiii  
List of Appendices ................................................................................................................ xiv  
List of Abbreviations, Symbols, and Nomenclature ............................................................... xv  
Chapter 1 ............................................................................................................................... 1  
1 General Introduction ........................................................................................................... 1  
1.1 The human microbiome .................................................................................................. 1  
1.2 The vaginal microbiome ............................................................................................... 2  
1.3 What is “normal”? ......................................................................................................... 3  
1.4 Conditions of the vaginal microbiome ......................................................................... 5  
  1.4.1 Bacterial vaginosis (BV) ......................................................................................... 5  
  1.4.2 Other microbiological disorders ............................................................................. 8  
1.5 Modulating the vaginal microbiome and treating aberrant conditions ..................... 11  
  1.5.1 Antimicrobials ..................................................................................................... 11  
  1.5.2 Probiotics ............................................................................................................ 12  
1.6 Other physiological states ............................................................................................. 15  
1.7 Key taxa and species-specific roles ............................................................................... 16  
1.8 High-throughput sequencing techniques for characterizing the microbiome ............. 19  
  1.8.1 Targeted amplicon sequencing (16S) .................................................................... 20  
  1.8.2 Genome sequencing, assembly, and annotation ................................................. 24
### Chapter 1

1.8.3 Metagenomics and metatranscriptomics ................................................................. 26  
1.9 Scope and objectives of the thesis .................................................................................. 29  
1.10 References ..................................................................................................................... 30  

### Chapter 2

2 Vaginal microbiome and epithelial gene expression of post-menopausal women with moderate to severe dryness ................................................................. 49  
2.1 Introduction ....................................................................................................................... 49  
2.2 Materials and Methods ..................................................................................................... 50  
  2.2.1 Study population ........................................................................................................... 50  
  2.2.2 Bacterial DNA extraction, amplification, and sequencing ............................................ 51  
  2.2.3 OTU clustering and taxonomic assignment .................................................................. 51  
  2.2.4 Vaginal epithelial RNA extraction .............................................................................. 52  
  2.2.5 Gene expression analysis .......................................................................................... 53  
  2.2.6 Statistical analyses ...................................................................................................... 53  
2.3 Results .............................................................................................................................. 56  
2.4 Discussion and Concluding Remarks ............................................................................... 61  
2.5 References ....................................................................................................................... 64  

### Chapter 3

3 At the crossroads of vaginal health and disease: the genome sequence of *Lactobacillus iners* AB-1 .............................................................................................................. 67  
3.1 Introduction ....................................................................................................................... 67  
3.2 Materials and Methods ..................................................................................................... 68  
  3.2.1 Organism ..................................................................................................................... 68  
  3.2.2 Genome assembly ......................................................................................................... 68  
  3.2.3 Gap closure and scaffolding ........................................................................................ 68  
  3.2.4 Validating genome coverage ...................................................................................... 70  
  3.2.5 Annotation .................................................................................................................... 70
4 Comparative meta-RNA-seq of the vaginal microbiota and differential expression by Lactobacillus iners in health and dysbiosis .................................................. 102

4.1 Introduction........................................................................................................... 102

4.2 Materials and Methods.......................................................................................... 103

4.2.1 Clinical samples.................................................................................................. 103

4.2.2 RNA isolation, mRNA enrichment, and sequencing......................................... 104

4.2.3 Reference sequence library and mapping ......................................................... 104

4.2.4 Functional assignment of refseqs....................................................................... 105

4.2.5 Statistical analyses for differential expression.................................................... 105

4.2.6 CRISPR spacer analysis....................................................................................... 105

4.2.7 cpn60 reference mapping..................................................................................... 106

4.3 Results and Discussion .......................................................................................... 106

4.3.1 Lactobacillus iners’ response to BV................................................................. 109

4.3.2 The vaginal community in BV............................................................................ 114

4.4 Concluding Remarks.............................................................................................. 116

4.5 References.............................................................................................................. 119

Chapter 5...................................................................................................................... 124

5 Effect of antimicrobial and probiotic therapy on the vaginal microbiota ............. 124

5.1 Introduction.............................................................................................................. 124

5.2 Materials and Methods.......................................................................................... 125

5.2.1 Clinical samples and study design..................................................................... 125

5.2.2 V6-targeted 16S rRNA gene sequencing.......................................................... 127

5.2.3 Sequence processing........................................................................................... 127

5.2.4 Data analysis ...................................................................................................... 127

5.3 Results...................................................................................................................... 128
5.3.1 The normal vaginal microbiota in pre-menopausal Brazilian women.... 128
5.3.2 Vulvovaginal candidiasis (VVC) study ........................................ 130
5.3.3 Bacterial vaginosis (BV) study .................................................. 133
5.4 Discussion .................................................................................. 136
5.5 Concluding Remarks .................................................................. 138
5.6 References ................................................................................ 138
Chapter 6 .......................................................................................... 142
6 General discussion ........................................................................ 142
  6.1 Defining normal ........................................................................ 142
  6.2 Bacterial vaginosis ..................................................................... 144
  6.3 The role of *Lactobacillus iners* ................................................. 147
  6.4 What are the bacteria doing? ..................................................... 149
  6.5 Considerations for modulating the ecosystem ............................ 153
  6.6 Future directions ....................................................................... 154
  6.7 References ................................................................................ 157
Appendices ......................................................................................... 162
List of Tables

Table 2-1. The top 20 genes up- and down- regulated genes ($p<0.05$) in the vaginal dryness group compared to controls. .......................................................... 59

Table 2-2. Enrichment of GO terms in the filtered gene list (n=960). ............................................. 60
List of Figures

Figure 1-1. The V6-sequenced vaginal bacterial profiles of 272 samples from 132 women. .. 9

Figure 1-2. “Omics” approaches for characterizing the human microbiome. .................... 20

Figure 1-3. Variability within the 16S rRNA gene................................................................. 21

Figure 1-4. Workflow used for multiplexed Illumina V6 sequencing for studies presented in
this thesis................................................................................................................................. 23

Figure 2-1. Microbiota profiles for 32 post-menopausal women clustered by biota similarity
.................................................................................................................................................. 54

Figure 2-2. Color taxa legend for Figure 2-1 and Figure 2-3 ................................................. 55

Figure 2-3. 16S (V6) microbiota profiles for 16 post-menopausal women sampled every 2
weeks........................................................................................................................................ 57

Figure 2-4. Heatmap of vaginal epithelial gene expression of 10 samples ......................... 58

Figure 3-1. Subset of a phylogenetic tree representing L. iners as part of the acidophilus
complex.................................................................................................................................... 76

Figure 3-2. Genomic atlas of Lactobacillus iners AB-1....................................................... 78

Figure 3-3. Venn diagram representing orthologous proteins between select species of the L.
iners clade ................................................................................................................................. 80

Figure 3-4. Comparison of the distribution of genes by COG functional category.......... 81

Figure 3-5. Alignment of conserved motifs in the putative cholesterol-dependent cytolysin
(LINAB1_0216) with related cytolysins ............................................................................. 84

Figure 3-6. Transmission electron micrograph of immunogold labeled L. iners AB-1 (black
arrows) in association with human vaginal epithelial cells ............................................ 89
Figure 4-1. Bacterial community structure and function ................................................................. 107

Figure 4-2. Circular representation of RNA-seq data for the *Lactobacillus iners* pan-transcriptome ......................................................................................................................................................................................... 110

Figure 4-3. Overview of predicted differential functions of the vaginal microbiota based on RNA-seq analysis ........................................................................................................................................................................................................................................... 118

Figure 5-1. Genus-level bar plot of vaginal bacterial abundance measured by V6 sequencing in 63 healthy, asymptomatic Brazilian women ................................................................................................................................................................................. 129

Figure 5-2. Distribution of Shannon diversity .................................................................................. 130

Figure 5-3. Genus-level bar plot of vaginal bacterial abundance measured by V6 sequencing in women diagnosed with vulvovaginal candidiasis (VVC) ........................................................................................................................................................................... 131

Figure 5-4. Paired UniFrac distance between individual patients before and after treatment for BV or VVC ........................................................................................................................................................................................................................................... 132

Figure 5-5. Violin plots showing the change in bacterial abundance within patients before and after treatment ........................................................................................................................................................................................................................................... 133

Figure 5-6. Genus-level bar plot of vaginal bacterial abundance measured by V6 sequencing in women diagnosed with bacterial vaginosis (BV) ........................................................................................................................................................................................................................................... 134

Figure 6-1. Summary of mRNA and V6 16S rRNA gene sequence data for 28 vaginal samples ........................................................................................................................................................................................................................................... 152
List of Appendices

Appendix A: Copyright from PLoS ONE ................................................................. 163
Appendix B: Copyright from PNAS ................................................................. 164
Appendix C: Copyright from Microbiome/BioMed Central ........................................ 165
Appendix D: Ethics approval for study presented in Chapter 2 .................................. 166
Appendix E: Ethics approval for study presented in Chapter 3 .................................. 167
Appendix F: Ethics approval for study presented in Chapter 4 .................................. 168
Appendix G: Ethics approval for study presented in Chapter 5 .................................. 169
Appendix H: Ethics approval for study presented in Chapter 6 .................................. 170
List of Abbreviations, Symbols, and Nomenclature

ALDEx         ANOVA-like differential expression (a software package implemented in R)
BLAST         Basic local alignment search tool
bp            Base pair
BV            Bacterial vaginosis
BVAB          Bacterial vaginosis-associated bacterium
CAI           Codon adaptation index
Cas           CRISPR-associated
CDC           Cholesterol-dependent cytolysin
CDS           Coding DNA sequence
COG           Clusters of orthologous groups of proteins
CRISPR        Clustered regularly interspaced short palindromic repeats
DNA           Deoxyribonucleic acid
GIT           Gastro-intestinal tract
GL            Glycerolipid
GLP           Glycerophospholipid
Gly           Glycerol
Gly-3P         Glycerol-3-phosphate
GO             Gene Ontology
HIV           Human immunodeficiency virus
HTG           Horizontally transferred genes
KEGG          Kyoto encyclopedia of genes and genomes
mRNA          Messenger ribonucleic acid
NCBI          National Center for Biotechnology Information
nrdb          Non-redundant database
nt            Nucleotide
ORF           Open reading frame
OTU           Operational taxonomic unit
PCR           Polymerase chain reaction
PTS           Phosphotransferase transport system
RAST          Rapid annotation using subsystem technology
RDP           Ribosomal database project
refseq        Reference sequence: referring to representative sequence in a cluster
RefSeq        NCBI reference sequence database
RNA           Ribonucleic acid
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>rRNA</td>
<td>Ribosomal ribonucleic acid</td>
</tr>
<tr>
<td>SDP</td>
<td>Sortase-dependent protein</td>
</tr>
<tr>
<td>STI</td>
<td>Sexually transmitted infection</td>
</tr>
<tr>
<td>subsys</td>
<td>Subsystem (referring to SEED level subsystems)</td>
</tr>
<tr>
<td>VVA</td>
<td>Vulvovaginal atrophy</td>
</tr>
<tr>
<td>VVC</td>
<td>Vulvovaginal candidiasis</td>
</tr>
</tbody>
</table>
Chapter 1

1 General Introduction

Sometimes referred to as our “second genome”, the human microbiome consists of all the microorganisms living on and in our human body. The number of microbial cells outnumber our own somatic cells 10 to 1 (1) and the genomic potential of these organisms is estimated to be 30X our own genome (2). These microbes exert an enormous influence on their hosts via the production of metabolic end products, acting as a line of defense against exogenous pathogens, and modulating host immune function. In the human vagina, the resident microbes reside in and on the surface of the vaginal epithelium relying primarily on host products for nutrients and in turn establishing a first line of defense against exogenous disturbances.

The broad goal of my thesis was to characterize the vaginal microbiome under healthy and altered states using high-throughput sequencing techniques. I also sought to describe the genome and transcriptome of Lactobacillus iners (Chapter 3 and Chapter 4) - the most prevalent organism of the human vaginal microbiome. Over the course of my thesis I helped develop new computational pipelines and tools for targeted 16S rRNA gene sequencing (Chapter 1 and Chapter 1), genome annotation (Chapter 3), and statistical evaluation of differential expression using mRNAseq (Chapter 4). We used these methodologies to show how the vaginal microbiota adapts upon antimicrobial and probiotic therapy of bacterial vaginosis (BV) and vulvovaginal candidiasis (VVC) (Chapter 5).

1.1 The human microbiome

Although consisting of prokaryotic, archaea, viral, and eukaryotic organisms, the “human microbiome” colloquially refers mainly to the bacterial populations detected in association with the human body as these are the most easily detectable organisms. Many of these bacterial communities are thought to have a commensal or mutualistic relationship with the host. For example, organisms in the gastro-intestinal tract (GIT) receive nutrients primarily from the host ingesting food, while in return maintain the
immune environment of the GIT, counter pathogenesis (3), and release nutrients from otherwise indigestible components of our diet (4). Some organisms may have pathogenic potential, but it is typically thought that a healthy microbiome maintains an ecological balance whereby the pathogenic organisms are kept under control by the rest of the population. Fluctuations in abundance or structure of these populations are common but typically the system returns to homeostasis in healthy individuals (5). A large and sustained disturbance can result in, or be associated with, a dysbiotic state and lead to conditions or diseases associated with changes in the normal bacterial ecology. Examples include inflammatory bowel disease (6), dental caries (7), and obesity (5).

Our current understanding of the human microbiome has largely been due to the advancement in sequencing technologies allowing for high-throughput assessment of microbiome population structures (the sequencing techniques are reviewed in Section 1.8). These studies have shown different body sites have distinct microbiological communities, and these communities differ between individuals (8).

1.2 The vaginal microbiome

From birth and over a woman’s lifetime there are changes associated with the vaginal physiology and correspondingly with the vaginal microbiome. From birth, babies born via vaginal delivery have biota of the skin, mouth, and gut that resemble their mother’s vaginal microbiome (9). As the infant grows, the different body sites acquire unique communities that increase the overall diversity of bacteria and differentiate the sites by their composition (10). Before puberty, the vaginal microbiome is dominated by anaerobes (11), after which estrogen levels increase, the vaginal lining responds with increased glycogen production, and the vagina of most women is colonized by species of Lactobacillus (12). The lactobacilli and other lactic acid bacteria convert the vaginal glycogen to lactic acid and acidify the vagina to a pH ≤4.5 (13). This balance is typically maintained until menopause when a decrease in estrogen results in a loss of glycogen and a loss of the lactobacilli (14).

Culture-based techniques dominated microbiology for many decades and a superficial understanding of the vaginal bacterial community was gained. Although isolation of
bacteria allows researchers to assay their biochemical abilities, most of the species that make up the human microbiome have not yet been cultured due to strict requirements that we cannot replicate in vitro (15, 16). The culture bias has carried over to the vaginal microbiota; for example, *Lactobacillus iners* is now known to be a major constituent of the vaginal microbiota, but it wasn’t isolated or described until relatively recently (17). Currently, there are several “bacterial vaginosis-associated bacteria” (BVAB) of which only one member has been isolated, and none have been fully described taxonomically, yet are detectable by molecular techniques (18, 19).

1.3 What is “normal”?

It’s theorized that the vagina is maintained at a slightly acidic pH to prevent the growth of harmful microorganisms (20). For most women, the production of lactic acid is driven by members of the *Lactobacillus* genus which dominate during healthy conditions. Vaginal lactobacilli were first described by Albert Döderlein in 1892 and their presence in discharge was quickly accepted as part of the normal status (21). Later studies revealed there were several related but distinct species of *Lactobacillus* in the human vagina including *L. jensenii*, *L. crispatus*, *L. gasseri*, *L. fermentum*, and *L. reuteri* (22, 23). Although the idea of a healthy *Lactobacillus*-dominated biota prevailed for some time, more recent molecular descriptions have expanded our knowledge of the species present and have indicated that the vaginal biota of up to 20% of women deemed to be healthy are not dominated by lactobacilli but have a more complex microbiota and an associated higher pH (24). Still, a *Lactobacillus*-dominated vaginal microbiota seems to be a uniquely human trait as sequencing studies of non-human primates show them to be rare constituents of the vaginal microbiota (25). Instead, non-human primates have low glycogen, higher pH, and a vaginal biota generally resembling human BV (25).

Early studies using 16S ribotyping still relied on isolating strains by culture before extracting their DNA and sequencing the 16S rRNA gene directly (26, 27), or by using gel separation(28). These revealed a diverse collection of species in the vagina that were not previously differentiated by biochemical assays and suggested that women were dominantly colonized by one species of *Lactobacillus*. Culture-independent 16S sequencing by direct isolation of total DNA from vaginal swabs revealed several species
previously rarely detected, like *Lactobacillus iners* (29), *Atopobium vaginae*, *Megasphaera*, *Leptotrichia* (30), *Bifidobacterium*, *Prevotella*, and *Streptococcus* (31, 32) as major constituents of the vaginal microbiome.

When technological advances allowed 16S sequencing to become high-throughput, the complexity of the vaginal microbiome was revealed. Papers by Sundquist *et al.* (33) and Spear *et al.* (34) have outlined some of the first approaches to distinguishing bacterial species in the vagina using partial 16S sequence and high-throughput pyrosequencing. Compared to other body sites like the gut and mouth, fewer species were detected and a healthy vaginal environment is now generally associated with a low diversity community (35). For other sites, a highly diversified community is often associated with health, and a loss in diversity is correlated with several conditions. The largest study to date of the healthy vaginal microbiome used partial 16S sequencing to determine the vaginal profiles of 396 reproductive-aged women of various self-reported ethnicities (white, black, Asian, and Hispanic) amongst Americans (24). The results suggested vaginal profiles fell into 5 distinct groups based on the organism dominance: four *Lactobacillus* groups dominated by one species (*L. iners, L. crispatus, L. gasseri, or L. jensenii*) and a fifth with a high proportion of mixed anaerobic bacteria. Interestingly, the distribution of the groups by ethnicities was not equal and a *Lactobacillus*-dominated profile was more likely for white and Asian women, while black and Hispanic more often had mixed communities. This also corresponded to an increased vaginal pH measured in black and Hispanic women compared to white and Asian women. These results support an earlier study (36) that concluded black women have a higher vaginal pH than other ethnicities. This possibly explains the apparent higher prevalence of ‘asymptomatic’ BV in African or black women (37). Another partial 16S rRNA gene sequencing study targeting 220 women with and without BV (as assessed by Amsel’s criteria which is based upon pH, discharge, and odor) also found women without signs of BV were dominated by either *L. iners* or *L. crispatus* with other *Lactobacillus* species co-existing in smaller proportions (38). Data from the Vaginal Human Microbiome Project at Virginia Commonwealth University has also shown the majority of women without obvious signs or symptoms of abnormal vaginal conditions are dominated with *Lactobacillus*, but the remaining show greater
diversity with a more even mixture of species (39). This implies that low diversity in the vagina is associated with health, contrary to the GIT and mouth.

There are few longitudinal studies of the vaginal microbiome, and little is known about the short-term changes in the community structure. The most extensive, to date, by Gajer et al. (40) sampled 32 women over a 16 week period and used partial 16S rRNA gene sequencing to track the changes in community composition. Notably, there was a lot of variation reported: some women had very stable communities, while others fluctuated rapidly over the study period. The authors theorized that most of the changes to community structure are tolerable because the functions, such as lactic acid production, are maintained. However, there were also clear perturbations to the environment represented by changes in detectable metabolic products and in response to menses. In another study using quantitative polymerase chain reaction (qPCR), Srinivasan et al. (19) described fluctuations of bacterial species during menses with an increase of L. iners and G. vaginalis. Such environmental perturbations may therefore lead to changes in the community structure and possibly allow pathogenic, or disease-associate organisms to establish.

As a consequence of achieving a high-resolution molecular picture of the vaginal microbiome via high-throughput sequencing, the point of view shifted from organism-centric to an ecological perspective, and the next step was to understand the interactions between all the organisms, environment, and the host’s response.

1.4 Conditions of the vaginal microbiome

1.4.1 Bacterial vaginosis (BV)

Bacterial vaginosis is the most common vaginal condition causing women to visit their doctors. Recent estimates have suggested 1 in 3 women will experience BV in their lifetime and the prevalence is up to 50% in some populations (41). First referred to as “non-specific vaginitis” early cases were described as having abundant discharge and a “fishy” odor not caused by infectious Candida or Trichomonas, and with a causative agent of Haemophilus vaginalis (later, Gardnerella vaginalis) (42). Examinations of vaginal smears and cultures in the early 1990s indicated there were notably fewer
lactobacilli and an overgrowth, sometimes in excess of 10 to 100-fold increase, of anaerobic species (43). However, the isolation of *G. vaginalis* from ~50% of those diagnosed with BV (44) from women without any signs of vaginal abnormalities (45) meant that *G. vaginalis* does not fulfill the criteria of Koch’s postulates (46) for causal relationship to BV. Indeed, despite many years of study, no single organism has been indicated as the causative factor in BV (47), and recent evidence has shifted the paradigm from an infectious disease to a polymicrobial dysbiosis.

The molecular descriptions of the BV microbiota using 16S rRNA gene sequencing have shown there are multiple diverse communities represented in this condition (18, 38, 48, 49). Most often, these communities show an increased abundance of anaerobic bacteria with a loss of lactobacilli. The predominant genera detected during BV include *Prevotella*, *Mobiluncus*, *Atopobium*, *Clostridiales*, *Leptotrichia*, *Sneathia*, and *Megasphaera*. Some studies have attempted to define bacterial clusters or groups of BV based on organism abundance and presence across individuals (38, 48-50) (and Figure 1-1). However, unlike healthy profiles that can generally be divided into distinct subtypes based on the dominating *Lactobacillus* species, BV clusters are not well defined due to the higher diversity and the difficulty in comparing data between studies. Whether different BV subtypes represent different relative risks for the host or not is still unclear.

BV affects the quality of life for women and their partners and presents an emotional and physical burden (51, 52). It is associated with more serious complications including an increased rate of HIV and other STIs (53), pelvic inflammatory disease, cervicitis (54), and higher incidence of pre-term labour (55).

The triggers that shift the microbiota to BV are not well understood, but the biggest risk factors include a past history of the condition (56), having multiple male sex partners or having female sex partners (57), and douching (58).

Clinically and in the laboratory there are two primary methods used to diagnose BV: the Amsel criteria as mentioned previously (44) and Nugent scoring (59). The Amsel criteria uses 3 of 4 of the following results to confirm BV: 1) Thin, white, homogenous discharge, 2) Clue cells (epithelial cells covered in bacteria) under microscopic
examination of a vaginal swab smear, 3) Vaginal fluid pH >4.5, 4) Positive “whiff” test: release of a fishy odor after adding 10% potassium hydroxide to vaginal fluid. The Nugent score is assessed from a vaginal smear viewed under 100X oil immersion microscopy. After Gram staining, the cell morphologies are counted under several fields of view. Gram-positive rods are presumptive \textit{Lactobacillus} and are scored low, while small Gram-variable rods (presumptive \textit{Gardnerella}), and curved Gram-variable rods (presumptive \textit{Mobiluncus}) are scored higher. The final score ranges from 0-10 with 0-3 a “Normal” score, 4-6 “Intermediate”, and 7-10 as BV. Variations and alternatives to the Nugent or Amsel criteria are sometimes used to assess vaginal microbiota health clinically, such as Hays method (60), or wet mount microscopy (61). Other kits, such as the BV Blue (62) and FemExam (63) have been tried with some success but not yet widely utilized, perhaps due to cost.

Evaluating the effectiveness of BV diagnosis methods is difficult considering the lack of consensus of what BV actually is, and whether signs (such as discharge and odor) are required for the condition to exist, or if a change in microbiological profile regardless of symptomatic presentation is indicative of BV. Additionally, there is weak correlation between Amsel and Nugent criteria (38, 64, 65) - the two most widely used diagnostic methods. Srinivasan et al. (38) attempted to correlate individual taxa defined by partial 16S rRNA gene sequencing with the four components of the Amsel criteria. Only \textit{Leptotrichia amnionii} and \textit{Eggerthella} sp. were correlated with all four criteria, and \textit{G. vaginalis} and \textit{A. vaginae} were associated with 3 of 4 criteria each. The other criteria had various taxa associations suggesting that different bacterial components could impact the symptomology of BV, which might explain some of the disassociation between Nugent and Amsel diagnoses.

Current clinical practices and diagnostic methods have not caught up with the molecular understanding of the diversity of vaginal microbiota. This has led to an association fallacy: although BV is nearly always associated with increased bacterial diversity, not all increases in bacterial diversity are associated with BV. Similarly, not all asymptomatic BV cases have \textit{Lactobacillus}-dominated vaginal microbiota, but nearly all women with \textit{Lactobacillus}-dominated vaginal microbiota are asymptomatic.


1.4.2 Other microbiological disorders

BV is the most recognized and common vaginal microbiota disorder, but there is a lesser-known bacterial dysbiotic condition known as aerobic vaginitis (AV) (66). Possibly due to lacking a routine microscopy based diagnostic testing in clinical settings, this condition is not often recognized. Like BV, AV is associated with a loss of lactobacilli, but instead of anaerobes the ecosystem has an overabundance of aerobic bacilli and cocci (such as *Streptococcus* species and *E. coli*). Unlike BV, AV is an inflammatory condition often occurring with tissue redness and inflammation, and presence of vaginal leukocytes visible under microscopic examination of vaginal smears (67). As expected, both conditions are deficient in lactate, but while BV is associated with end-products of anaerobic fermentation (mainly succinate), vaginal fluid from AV samples contain high levels of inflammatory cytokines interleukin (IL) 6, IL-8, and IL-1β (66). The medical implications of AV are not yet clear (68), although some association with preterm labour has been reported (68).

There are several infectious conditions of the vagina caused by specific organisms invariably acquired through sexual transmission. These include human immunodeficiency virus (HIV), human papilloma virus (HPV), herpes simplex virus (HSV), *Chlamydia trachomatis*, *Trichomonas vaginalis*, *Neisseria gonorrhoeae*, and others that can arise spontaneously like *Candida* (yeast), or group B *Streptococcus* infections. The vaginal mucosa and associated microbes provide the first line of defense against external or internal disturbances by other microorganisms. Thus when lactobacilli are displaced during BV, the risk for acquisition of several of these STIs increases (53).

HIV infection is of particular interest in relation to BV because of the overlapping high incidence in many developing countries, particularly in Africa (69). Studies uncovering the potential relationship between BV and HIV have shown a BV-type microbiota can increase viral shedding and viral load in the vagina (70, 71). A recent study suggested succinate, abundantly produced by BV-associated organisms, could increase HIV gene expression and inflammatory interleukin-8 production in infected differentiated monocyte-derived macrophages (72). Other studies show that BV itself increases the risk of acquisition of HIV (73).
Our recent study of the vaginal microbiota of 132 HIV-positive women in Tanzania (48) suggested that BV profiles, though diverse, could be grouped into major clusters represented by the dominating taxa (Figure 1-1). As supported by another study of HIV-positive women (34), we found that *L. iners* was the most commonly detected organism regardless of BV status, and that other *Lactobacillus* species were rare with the exception of one cluster of non-BV samples dominated by *L. crispatus*. The organisms associated with BV included *Lachnospiraceae* (later identified as BVAB1), *Prevotella*, *Megasphaera*, and *Veillonella* amongst others. These profiles are similar to those of BV+ HIV-negative women and so far there has been no evidence to suggest distinct BV profiles exist in HIV-positive women (74).

**Figure 1-1. The V6-sequenced vaginal bacterial profiles of 272 samples from 132 women.**

Each bar is a sample and each color is an OTU or representative taxon. Samples were clustered by the similarity in organism abundance as visualized by the neighbor-joining tree below the bar plot. The major clusters are identified above the bar plot by
the predominant organism(s). The sample numbers above the dendrogram are coloured according to Nugent categories with BV = red, intermediate = green, normal = blue. Amsel criteria are shown for each sample with present = grey, absent = white and missing data = black. Figure from (48).

After BV, vulvovaginal candidiasis (VVC) is the next most common vaginal microbial disorder (75). It is caused by Candida albicans in 70-90% of cases (75), but is also associated with C. glabrata (7-16% of cases), amongst others (76). It’s estimated that 70-75% of women will experience VVC in their lifetime (77). There is some evidence that the prevalence of VVC is higher in African-American women (78), and rare in post-menopausal women unless they are receiving estrogen therapy (77, 79). Although patients with BV and VVC present similar signs and symptoms of discharge and irritation, specific symptoms associated with VVC include inflammation, itching, pain during sex (dyspareunia) or urination (dysuria), and lack of odor and abnormally high pH associated with BV (80). Diagnosis is confirmed with positive selective Candida culture, and/or the presence of Candida by microscopy on a wet mount or Gram stain in conjunction with the associated signs and symptoms (80). However, vaginal Candida colonization is common with one study showing 15% of unselected asymptomatic women (81), and a larger longitudinal study reporting 70% of women were positive for Candida in at least one of four samples taken over the course of one year (82). The isolation of Candida therefore does not necessarily equate to infection.

The role of the bacterial microbiota in relation to VVC is largely unexplored. Antibiotic use is highly associated with incident of VVC with approximately 30% of treated women acquiring this condition after antibiotic use (83). The proposed mechanism of post-antibiotic VVC is thought to be due to disruption of the protective vaginal bacterial community thus allowing for Candida overgrowth. In vitro studies have suggested that Lactobacillus species common in the vagina could produce compounds that inhibit growth or adhesion of C. albicans providing potential mechanisms of exclusion (84, 85). One that tested a supposed Lactobacillus probiotic did not show any effect in preventing post-antibiotic VVC (86). However, this study does not indicate strain designations or reasons for selection, and included strains not typically found in abundance in the vagina (L. delbrueckii, L. acidophilus, Streptococcus thermophilus). Only a few studies have
examined the association of the bacterial vaginal microbiota in relation to VVC. One by Sobel and Chaim (87) found no evidence of differential *Lactobacillus* abundance in women with acute VVC compared to healthy asymptomatic women. The study found no evidence of differential *Lactobacillus* abundance between the groups. Similarly, Vitali *et al.* using PCR denaturing gradient gel electrophoresis (PCR-DGGE) and real-time PCR demonstrated both healthy conditions and VVC were dominated by *Lactobacillus* while only BV showed a significant deviation in lactobacilli (88). A study by Zhou *et al.* (89) used terminal restriction fragment length polymorphisms (T-RFLP) of the 16S rRNA gene, and supplemented with cloning of 16S rRNA gene fragments for sequencing to show no evidence of a significantly altered profile in women who hadn’t had a yeast infection in the past 2 years and those with a history of VVC with most women both groups dominated by *Lactobacillus* species. Unfortunately these studies are limited in determining the community composition due to culture bias, low specificity, and so far lack of use of high-throughput sequencing methods.

The function of the microbiome in other vaginal infections is not well understood. Only recently have studies emerged specifically examining the bacterial composition during *Trichomonas vaginalis* (90), *Chlamydia trachomatis* (91) and HPV (92, 93) infections. Considering the nearly 2-fold increased risk of acquiring these infections by having an abnormal BV-like profile (94, 95), and the interplay between host immunity and the vaginal microbiota (96) the composition and function of the vaginal microbiome is a key component to urogenital health and clearly needs to be investigated.

1.5 Modulating the vaginal microbiome and treating aberrant conditions

1.5.1 Antimicrobials

The current standard of treatment for BV is either an oral or intravaginal application of metronidazole or clindamycin daily or twice a day for approximately 1 week (97). These antibiotic therapies have been unchanged for over two decades despite the high recurrence rate of BV. Up to 50% or more women have a recurrent BV episode within 12 months of standard treatment (56), although the reasons are not known. Because of the
associated risks of BV, especially for complications during pregnancy, many women without symptoms but who show an abnormal Gram smear may be treated with antibiotics (98). However, this treatment, especially using metronidazole, has so far failed to prevent pre-term labour (98, 99). For VVC, topical or oral anti-fungals such as fluconazole are effective in eradicating Candida overgrowth and eliminate symptoms in 90% of uncomplicated VVC cases (77), perhaps because of BV biofilms, poor specificity or rapid recovery of the pathogens (100).

For VVC, topical or oral anti-fungals such as fluconazole are effective in eradicating Candida overgrowth and eliminating symptoms in 90% of uncomplicated VVC cases (77). VVC caused by non-albicans species are more likely to be resistant to standard treatment and are more associated with recurrent VVC (81).

Because antibiotic usage has side effects including nausea, vomiting, diarrhea, and risks like thrombophlebitis, gastrointestinal upset, development of resistant organisms, and recurrent episodes, and in the case of BV especially, the general ineffectiveness of the application, there is an urgent need for more robust treatment. This will be optimized the more we understand what is actually happening in the microbiome. With no sign of new pharmaceutical agents being developed, other options have been explored by scientists, one of which has been probiotics.

1.5.2 Probiotics

Probiotics, literally meaning “for life”, is currently defined by the FAO/WHO as: "live microorganisms which when administered in adequate amounts confer a health benefit on the host” (101). Probiotics consists of single strains or mixtures administered via oral or topical application with the intended goal of effecting the site of action in one or more ways by: 1) displacing pathogens, 2) modulating local or systemic immunity, 3) altering the microbiota community directly or indirectly. The means these outcomes are accomplished can be multifaceted and include: adhesion by the probiotic organism to host mucosa, up-taking or releasing compounds and metabolic products, surface interaction with the host immune system, or direct competition with resident microbes (102). The most common probiotics for urogenital health are strains of Lactobacillus
administered vaginally or orally. Strains have been selected based upon their ability to promote lactobacilli colonization, inhibit pathogen growth and adhesion and modulate host immunity (85, 103-105, 105-107). Several clinical studies have assessed the effects of *Lactobacillus* probiotic for treatment or prevention of BV alone or in conjunction with antibiotic therapy (reviewed by (108-110)). These have used *L. acidophilus* (111), *L. delbrueckii* subsp. *lactis* DM8909 (112), *L. brevis* CD2 + *L. salivarius* FV2 + *L. plantarum* FV9 (113), and *L. rhamnosus* Lcr35 (114), and some have been used to prevent recurrent BV like *L. gasseri* (Lba EB01-DSM 14869) + *L. rhamnosus* (Lbp PB01-DSM 14870) (115) and *L. rhamnosus* + *L. acidophilus*, + *Streptococcus thermophilus* (116). Unfortunately, many of these studies are limited in sample size, trial design, strain selection and reporting, and choice of dosage and mode of administration. However, data from well-controlled, randomized trials suggest some evidence for beneficial effects of administered lactobacilli for vaginal health – in particular, the combination of *L. reuteri* RC-14 and *L. rhamnosus* GR-1 have been effective for treatment (63, 117-119) and prophylaxis (118, 120) of BV.

The potential probiotic mechanisms of these strains have been well characterized. *L. rhamnosus* GR-1 adheres to uroepithelial cells, and can effectively exclude and inhibit growth of uropathogens (103). Along with the ability to adhere to uroepithelial cells (121), *L. reuteri* RC-14 is able to produce abundant hydrogen peroxide (122), and a biosurfactant that effectively inhibits adhesion by uropathogenic bacteria (123), and can disrupt *G. vaginalis* biofilms *in vitro* (124, 125). Both RC-14 and GR-1 strains have been shown to persist in the vaginal tract for up to 19 days post intravaginal application for 3 days, while the probiotic gastrointestinal isolate *L. rhamnosus* GG was not detectable after 5 days (126). Reid *et al.* (127) were the first to show recovery from the vagina of orally administered *L. rhamnosus* GR-1 and *L. reuteri* RC-14; later independently confirmed (128, 129). Studies showed these orally administered organisms could alter the vaginal biota in healthy reproductive-aged women and post-menopausal women by increasing the lactobacilli load (63, 119, 130, 131). The reported effects of this combination probiotic treatment in most cases is approximately twice the BV cure rate compared to antibiotic therapy (117, 120, 132).
For VVC, the high prevalence and number of women seeking relief from recurrent or post-antibiotic VVC have led to over-the-counter antifungal medications, often incorrectly administered when the patient actually has BV or another condition (133) and seeking treatment with probiotics. Approximately 40% of women reporting post-antibiotic VVC have used probiotic or yogurt products for prevention or treatment. This is second only to over-the-counter antifungal usage for treatment (63%) or prevention (49%) of VVC (134). Of note, many women apparently use probiotics to treat VVC, when there is little evidence to suggest this is efficacious on its own. Few studies have actually assessed the effect of probiotic lactobacilli on VVC possibly due to lacking evidence that women with VVC have an altered vaginal biota as in the case of BV (77). The combination strains \textit{L. reuteri} RC-14 and \textit{L. rhamnosus} GR-1 with fluconazole improved the cure rate of VVC compared to fluconazole alone (135, 136). Chapter 5 presents a follow-up to the Martinez study by examining the change in the vaginal microbiota following standard and probiotic-augmented treatment. Studies of potential mechanisms show that the combination strains can directly inhibit and kill \textit{Candida albicans} (137), and RC-14 can reduce IL-8 and IP-10 secretion by VK2/E6E7 cells induced by \textit{C. albicans} (138). Another study showed that these two lactobacilli suppress expression of NF-κB-related inflammatory genes, and may induce IL-1α and IL-1β expression by an alternate signal transduction pathway, such as MAPK/AP-1 (139). Still, with insufficient sample sized clinical data, it is too soon to make conclusions about the effectiveness of probiotic therapy for VVC (77).

Probiotics have had varying success for treatment of other vaginal conditions. Post-menopausal women represent a cohort deficient in vaginal lactobacilli compared to reproductive-aged women. In a double-blind, placebo-controlled study Petrichevic \textit{et al.} used orally administered \textit{L. reuteri} RC-14 with \textit{L. rhamnosus} GR-1 to restore \textit{Lactobacillus} in the vagina resulting in a normalized Nugent score (131). The first evidence for intravaginal use of probiotics to reduce recurrences of urinary tract infection (UTI) came from \textit{L. rhamnosus} GR-1 alone and with \textit{L. fermentum} B-54 (140, 141). The failure of other \textit{Lactobacillus}-based probiotics (109) demonstrates the importance of strain characteristics. A strain of \textit{Lactobacillus crispatus} CTV-05 has been tested in an early phase study to prevent urinary tract infections (UTIs) by repopulating the vagina
In general, studies using probiotics for treatment of urogenital conditions are successful when the strains are selected from the urogenital tract, and have properties allowing them to compete with the vaginal pathogens (121, 143). In the case of BV especially, probiotic lactobacilli can aid in restoring a *Lactobacillus*-dominated biota. Overall, probiotics are a promising therapy for vaginal and bladder health and understanding the mechanisms of actions and the interplay between host genetics, immunity, and natural microbiota will aid in designing or selecting strains with the most benefit for the individual condition.

### 1.6 Other physiological states

Hormone changes throughout a woman’s lifespan are associated with physiological responses that can affect the vaginal microbiota. Puberty, pregnancy and menopause are the biggest events causing hormone fluctuations. During pregnancy an influx of estrogen and progesterone from 50 to 1000 times pre-pregnancy levels (144) increases the glycogen content of the vaginal epithelia (12). Although the vaginal microbiota during pregnancy is largely unexplored, a recent 16S sequencing study by Aagaard *et al.* (145) suggests that compared to non-pregnant women there is a lower diversity of organisms with a dominance of *Lactobacillus* species. In preliminary studies, this has not been our experience when analyzing samples with the Illumina platform (unpublished data). Dominguez-Bello *et al.* (9) demonstrated that babies delivered by vaginal birth had whole-body biotas similar to their mother’s vaginal biota and dominated by lactobacilli, while babies born by cesarean section had biotas made up of predominantly skin bacteria such as *Staphylococcus, Corynebacterium,* and *Propionibacterium*. Babies born by C-section have a greater incidence of allergies and asthma, and of methicillin-resistant *S. aureus* infection after birth (9). Thus the lactobacilli may be providing a protective role in pathogen exclusion and immune mediation in newborns, and changes in the vagina during pregnancy could theoretically prime the environment for establishment of lactobacilli before birth. Supporting this, a disrupted vaginal microbiota in the form of BV is associated with pregnancy complications leading to risk of pre-term labour (146), and several organisms found in the vagina, such as *Ureaplasma, Dialister, Sneathia,*
Gardnerella), have been detected in amniotic fluid after complications such as intra-amniotic inflammation (147).

Conversely to pregnancy, menopause is marked by a decreased production of estradiol and progesterone by the ovaries. The resulting 7-fold drop in plasma circulating estrogen (148) has effects on the vagina including thinning of the membranes, decreased blood flow, and loss of elasticity leading to vaginal dryness and atrophy (149). Associated with these changes is a loss of lactobacilli after menopause (14, 150, 151) and elevated vaginal pH above the normal 4.5 to 6.0-7.5 (152). In general, the lactobacilli depletion does not result in colonization by BV-associated organisms (14, 151), and the Nugent scoring system is unreliable to determine BV status post-menopause (14, 153). Indeed, when these women are colonized by large numbers of lactobacilli they have lower frequencies of vaginal E. coli – an organism significantly associated with UTI (154). Hormone replacement therapy is often administered to alleviate adverse symptoms of menopause, vaginal dryness (a result of atrophy and decreased vaginal secretions), and dyspareunia (painful sexual intercourse), by re-introducing estrogen, progesterone, and/or progestin orally or vaginally. As a result lactobacilli are restored to the vaginal microbiota (150, 153, 155) and vaginal pH decreases to a normal 4.5 (156), and incidence of UTI are reduced (157).

1.7 Key taxa and species-specific roles

Species of Lactobacillus have long been associated with the human vaginal microbiota, though they are found in smaller proportions in the gastrointestinal tract and oral cavity (158). The genus is a member of the lactic acid bacteria group: a clade of Gram-positive, low-GC bacteria that produce lactic acid as a major fermentation product. This characteristic is thought to have a major functional role in the vagina where abundant lactic acid maintains a low pH unfavorable to pathogenic organisms (159, 160). This represents a potential mutualistic relationship where the vaginal epithelial cells produce glycogen that is converted by the bacteria to lactic acid (13). Lactobacillus species are also known to produce small antimicrobial molecules called bacteriocins, and hydrogen peroxide - both considered defense mechanisms against other organisms. Although recent studies have suggested H₂O₂ is not produced in sufficient quantities in the microaerobic
vaginal environment to have an antagonistic effect against other organisms (159, 160). Other than production of toxic metabolic byproducts, competitive exclusion by adherence to the mucosal surface and auto-aggregation has been proposed as an important protective strategy (161).

Revealed by comparative genomics, the genus *Lactobacillus* is highly diverse with evidence for extensive gene loss and gain via horizontal gene transfer (162). Species found in the vagina (including *L. crispatus, L. iners, L. gasseri, L. jensenii, L. johnsonii, L. vaginalis, L. reuteri*) are phylogenetically more similar than other *Lactobacillus* species (163, 164).

Notably, although BV is in general a low-*Lactobacillus* state, several studies have indicated *L. iners* can remain in relatively high abundance during BV (18, 48, 165) and post-antimicrobial treatment (166). This species was overlooked due to non-standard cultivation requirements until it was cultured on blood agar and described by Falsen *et al.* (17). Molecular sequencing techniques have uncovered *L. iners* as the most common constituent of the vaginal microbiota (24, 29, 48). Unlike other *Lactobacillus* species, *L. iners* seems to exclusively inhabit the human vagina and isn’t detectable in saliva or fecal samples (167), nor has it been reported in non-human primate vaginas (25). Although notably, the methods used in these studies may be limited in detection sensitivity.

Genome sequencing and *in vitro* assays have uncovered functions of particular strains that could explain their role in the vagina during BV. *Gardnerella vaginalis* is abundant in many cases of BV, but is a common constituent of the normal vaginal biota (45). This observation has lead to studies that suggest different strains of *G. vaginalis* could have different functions and result in varied risks for the host (168). Genome studies of *G. vaginalis*, once thought to be a single organism, reveal a diverse complement of genes between different strains with as much as 31% divergence in gene content and order (169). Two studies have used comparative genomic analysis on strains isolated from healthy or asymptomatic women compared to BV-isolated strains. Harwich *et al.* (170) found a point mutation in the protein encoding vaginolysin, a cytolysin active against human cells (171), that rendered it less cytotoxic than the same protein from the health-
associated strain. The same BV-associated strain had better adherence, aggregation, and biofilm formation. Biofilm formation is thought to be one of the major strategies *G. vaginalis* uses to persist (172) and resist antimicrobial compounds (173) and lactic acid (174). In addition, adherence of *G. vaginalis* biofilms to vaginal epithelial cells is responsible for the “clue cells” observed during BV and used as part of the Amsel diagnostic (175). The second comparative genomic study of four strains by Yeoman *et al.* (176) found a number of conserved genes with virulent potential including exopolysaccharide (EPS) biosynthesis for biofilm formation, pili, hemolytic/cytolytic potential, and several bactericidal toxins. Of the four strains, one (409-05) associated with a healthy asymptomatic woman was lacking genes encoding enzymes for mucin degradation – a trait associated with BV (177). In comparison to other BV isolates from different genera, *G. vaginalis* strains were determined to have a higher virulent potential by Patterson *et al.* (178) by assaying adherence, cytotoxicity, and biofilm activity. One study has implicated *G. vaginalis* in a symbiotic relationship with another BV-associated organism *Prevotella bivia* whereby amino acids produced by *G. vaginalis* supports the growth of *P. bivia* that in turn produces ammonia utilized by *G. vaginalis* (179). A similar symbiosis has been described between *P. bivia* and *Peptostreptococcus anaerobius* (180). *Prevotella* and *Gardnerella* species, and in particular *P. bivia*, have been shown to have sialidase activity (181) which is associated with risk of pre-term labour (182).

Other bacteria associated with the vaginal microbiota are less characterized due to the effort required to culture them for biochemical assays and sequencing (183). *Atopobium vaginae* has only been described relatively recently (184), but has a high specificity with BV where it is detectible by sequencing methods (183, 185). Most strains are resistant to metronidazole (186), and at least one affects local immunity by inducing release of IL-6, IL-8, and β-defensin 4 from epithelial cells in vitro (187). Other recently detected organisms include the BV-associated bacteria (BVAB) of the *Clostridiales* order which have high specificity for BV (18). Many of these low abundance, difficult-to-culture species, could be forming co-occurrence networks with other species in the vagina, as suggested by Srinivasan *et al.* (38). The collective effort of the individual strains could benefit the community as a whole if, for example, several mucinases produced by
different members of the community allowed access to the vaginal epithelium where attachment and lysis can occur (46). These yet to be characterized organisms reflect the complexity of the vaginal environment and the potential interactions that can occur between multiple taxa contributing to the etiology of BV and other conditions.

1.8 High-throughput sequencing techniques for characterizing the microbiome

Recent years have seen an upsurge in sequencing techniques to characterize the microbes of various environments including the human microbiome. Thanks to the development of several “next-generation” technologies including 454 pyrosequencing (Roche), Illumina sequencing by synthesis (previously Solexa), and later SOLiD (ABI), and Ion Torrent (Life Technologies) the cost of sequencing has dropped and the throughput has risen dramatically – far outpacing the predicted rate of technological development by Moore’s Law (188). These sequencing technologies produce millions of short sequence fragments (reads) in a single run. This allows for multiplexing of samples (multiple samples can be sequenced on the same run) and enables a greater sequencing depth (more reads per sample) allowing detection of organisms in lower abundance. There are a few main strategies, described here and illustrated in Figure 1-2, which give different information about the community structure or function. Many of the biggest challenges lay in the computational analyses of these data that are highly diverse and multivariate.
Figure 1-2. “Omics” approaches for characterizing the human microbiome.

The scope of this thesis uses next-generation sequencing and bioinformatics to describe the vaginal microbiome and the DNA and RNA level. Further information can be gained by examining the protein and metabolite populations in association to the microbiota.

1.8.1 Targeted amplicon sequencing (16S)

Currently the most common technique for characterizing a bacterial community is targeted 16S rRNA gene sequencing for molecular identification of bacterial taxa and their relative abundance in the sample. Using targeted PCR and high-throughput pooled sequencing, the relative abundance of assigned taxa can be determined for 100s or even 1000s of samples at a time. The 16S rRNA gene is conserved in all bacterial species and forms part of the small ribosomal subunit. Woese and Fox (189) were the first to use 16S rRNA gene sequence to infer phylogenetic relationships between different organisms, a technique that is now pervasive in microbiology. The gene is well suited for targeted PCR because of several so-called “variable” regions flanked by more conserved sequence (Figure 1-3). These stretches of variable sequence often have enough differences between
organisms to distinguish genus, species, and sometimes strain taxonomic levels using sequencing information alone (190).

**Figure 1-3. Variability within the 16S rRNA gene**

From pre-aligned sequenced >1200 bp downloaded from RDP, the variability, measured as Shannon information entropy, was calculated at each sequence position, using only positions without a gap in *E. coli*. The graph shows the Shannon entropy (y-axis) averaged over 50 bp windows, centered at each position in the gene (x-axis). Shannon entropy at position *x* was calculated as $-\sum p(x_i) \log_2 p(x_i)$, where $p(x_i)$ denotes the frequency of nucleotide *i*. Variable regions are marked V1 to V9, and numbering on the x-axis corresponds to the *E. coli* sequence. The V6 region was the target amplimer for the studies presented in this thesis. Figure is adapted from Andersson *et al.* (191). © 2008 Andersson *et al.* This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

High-throughput sequencing removed much of throughput restrictions of clone- and gel-based sequencing methods, but at the cost of shorter sequence read lengths meaning only part of the 16S rRNA gene could be sequenced at a time. However, the majority of the taxonomic information from full-length 16S gene alignments can be captured with short fragments such as single variable regions (190), and sample clustering is robust to differing sequencing length and variable region (192). Though reads are short, high-
throughput sequencing allows for greater depth (more reads per sample) and thus more power for detecting rare ribotypes in the community.

The choice of variable region to sequence should depend on a number of factors: the length of the sequence read available, the complexity of the environment, and the number of closely related taxa in the environment. Though the 16S rRNA gene has become the gold standard, some groups have used other gene targets like *cpn60* (193), and *rpoB* (194). For our studies we determined the V6 region offered the most distinguishing power for *Lactobacillus* species in the vagina (Gloor, G.B., personal communication), as one of our goals was to differentiate species like *L. iners* and *L. crispatus*. The V6 region has the most sequence information content (Figure 1-3), and other studies have shown it to be informative for genus and species differentiation (195) although automated tools designed for longer variable regions do not perform well with V6 sequences (196). For the 16S vaginal profiles presented in Chapter 1 and Chapter 5 we developed a sequencing methodology including a primer set with dual barcodes for paired-end sequencing on the Illumina platform (197). As outlined in Figure 1-4, the V6 region of the 16S rRNA gene is amplified with a unique sequence barcode for each sample, and then amplicons from all samples are pooled for sequencing. Downstream bioinformatic analysis overlaps the paired reads into full-length V6 sequence, and uses the barcode sequence to assign reads back to samples. As is standard in the field, the complexity of the dataset is reduced by clustering reads using sequence identity into operational taxonomic units (OTUs). The OTUs are arbitrarily defined at a percent identity cutoff representing a unit of taxonomy. By convention, most researchers pick a 97% identity cutoff as a proxy for species separation (198). However, this is merely a convenient binning strategy as the sequence variation is not constant across the entire rRNA gene (Figure 1-3) and may not represent the overall dissimilarity between species, and within different taxonomic lineages the 3% difference does not hold true (199). A representative sequence for each OTU cluster, usually the most commonly occurring sequence, is chosen for taxonomic assignment. This strategy can reduce tens of millions of reads to hundreds or fewer while retaining information on relative abundance in a table of read counts per OTU per sample (197).
Taxonomic assignment uses sequence alignment to make presumptive classification of reads based on similarity to known sequences. This is accomplished using BLAST alignment (e.g. default Greengenes (200) classifier), global alignment (GAST) or by comparing small fragments of the read called k-mers to a database of sequences with defined k-mer frequencies (ribosomal database project (RDP) (201), Greengenes). Other than ensuring a good alignment, taxonomic assignment is only as good as the database. As a result of being able to identify more sequences by high-throughput sequencing, there is now a massive amount of sequence data in the databases that is not characterized or assigned to a taxonomic lineage. In response to this, several manually curated databases like Greengenes (200), RDP (201), and SILVA (202), have arisen in attempt to filter out unreliable uncultured and unclassified sequences. For vaginal organisms, the databases are underrepresented (203) making confident taxonomic assignments even more difficult.

At the community level there are two primary ways to evaluate diversity (204). Within a sample, the “alpha diversity” describes how many taxa are detected and their relative abundances. This can be calculated with a number of metrics, mostly borrowed from ecology, like Chao1 (205), Hill (206), Simpson’s (207), or Shannon’s (208) index. The “beta diversity” describes the overlapping taxa between samples, often accounting for differences in abundance as well as presence and absence. Based on taxa membership, a distance value is calculated for every sample against all other samples representing how
dissimilar one sample is to another. UniFrac is a popular distance metric developed for 16S sequence data that can use presence/absence (unweighted) or abundance information (weighted) to create the distance matrix of sample distances (209). UniFrac builds a phylogenetic tree based on OTU sequence alignment and the sample distance is calculated based on the branch lengths shared and unshared between samples. Alternatively, sequence-independent methods, like Bray-Curtis (210), can also be used to build a distance matrix. The relative OTU abundances and distance matrix can be combined with multivariate statistics, and visualization and clustering tools in order to describe the relationship of the samples in regards to their microbiota composition (211).

There are multiple tools (e.g. QIIME (212), Mothur (213), phyloseq (214) offering multiple strategies for 16S rRNA sequence analysis since the approach will depend on the type of data and goals of the study. In addition, the complexities of multivariate analysis for these data have not been conquered. For example, both alpha and beta diversity estimates make assumptions about the data and in certain cases, such as when there are abundant rare OTUs, the calculated diversity is not representative of reality (215).

Using 16S and other tag-sequencing approaches might address “who is there”, but the data do not disclose the functional role of the detected OTUs without severe inferences, and given that strain differences are particularly hard to distinguish with partial 16S sequencing, there is considerable genomic variation not being accounted for.

1.8.2 Genome sequencing, assembly, and annotation

If 16S rRNA gene sequencing addresses “who is there?” in a microbial community, then whole genome sequencing addresses “what can they do?”. Once prohibitively expensive and time-consuming, high-throughput sequencing has allowed sequencing of bacterial genomes to become routine. As an example of the rapid pace, as of 2006 (one year after Roche released its GS20 pyosequencer) there were 10 published Lactobacillus genome sequences and 11 more in progress (216). By September 2013 there are more than 290 Lactobacillus genome projects listed at the National Center for Biotechnology Information (NCBI).
Shotgun sequencing is the method of choice for most whole genome sequencing projects. Briefly, pure cultures of the strain of interest are grown and the total DNA is extracted from a large number of cells. The DNA is randomly fragmented, prepared as a sequence library for the platform of choice, and then sequenced. The resulting sequenced fragments can range from 35 to 800 nt in length, and are paired or single. Paired-end sequencing overcomes some of the limitations of short read length by instead partially sequencing both ends of a longer fragment of a known length. Since the input fragment length is known, the distance (in bp) between the sequenced paired ends is determinable and can be leveraged for downstream assembly. Single unpaired reads from pyrosequencing platforms are generally longer, but with the trade-off of fewer reads produced per run and higher error rates compared to shorter paired-end sequencing on the Illumina or SOLiD platforms (217).

De-novo genome assembly algorithms find overlapping sequence between reads and combines the overlapping reads to build longer stretches of contiguous sequence (contigs). If paired-end reads are used, the distances between the pairs can be incorporated into the assembly algorithm to place paired reads in contigs at the correct distance apart (218). Generally, there are checks and restrictions built into the algorithms to avoid misassembled sequence not representative of the originating genome. Reads are not usually assembled unless there are multiple reads representing the same sequence (high coverage), the paired ends are the correct distance apart when assembled, and the bases in the read sequence have been confidently called by the sequencer. Repetitive sequences in bacterial genomes present a challenge as the reads are not usually long enough to span the repetitive region causing breaks or gaps in the assembly when an unambiguous path cannot be resolved (219). If available, contigs from an assembled genome can be scaffolded (placed in chromosomal order) by alignment to another finished related genome. Evidence for placing contigs together in scaffolds can also come from paired reads (if pairs span two contigs), contextual clues from annotated genes known to be near each other in on the chromosome, or with the aid of PCR and optical mapping. As finishing and closing a genome is costly and time-consuming with little information gained compared to assembled contigs, the vast majority of bacterial genome projects do not go beyond the “draft” stage (220).
Once contigs or scaffolds are built, open reading frames and other genomic elements (non-coding RNA, promoters, operons, regulatory regions) can be annotated and the functions predicted. There are a large number of computational tools available for these tasks, but the challenge remains in making confident and well-supported functional predictions. Though time-consuming and labour-intensive, manually annotated genomes are considered better quality than automated strategies (220). As functional predictions are generally made by alignments to other known sequences from a database, errors are easily propagated and misannotations are common (221). Without biochemical and molecular experimental validation, the functional annotations can only be inferred. Nonetheless, bacterial genome sequencing has provided incredible insight into pathogenic function, phylogeny and diversity, genome structure, horizontal gene transfer and genome evolution, novel genes, and targets for vaccines (222).

In conjunction with high-throughput sequencing studies of the human microbiome, whole genomes are invaluable references allowing assignment of sequence information to specific organisms and functions. Whole genomes also provide contextual information of whole transcriptional units, operons, and cellular or metabolic networks for understanding how an organism functions.

1.8.3 Metagenomics and metatranscriptomics

One way to classify functions of a whole microbial community is to shotgun sequence all the genomic DNA (metagenomics) or RNA (metatranscriptomics) content from an environment. Like amplicon sequencing, metagenomic sequencing can be used to determine the bacterial abundances in the community, but with added information about gene function and abundance. There are fewer computational tools for metagenomic methods and the data analysis is more challenging (219). Short reads do not give as much information about functional assignment compared to full-length gene sequences where the predicted protein translation is generally more informative than nucleotide sequence (223). Translating and comparing hundreds of millions of small sequencing fragments is computationally intensive and time-consuming. To circumvent this, two main strategies are used: 1) mapping or binning data to known genomes or functional groups, 2) assembling the small fragments into larger contigs that can be better annotated. The first
case is reliant on the availability of sequenced and annotated genomes, while assembly is
dependent on the read length and complexity of the sequences. For both cases, with the
ultimate goal of assigning gene function, tools borrowed from genomic annotation can be
used, or more high-throughput options such as MEGAN (224) or MG-RAST (225).

Metatranscriptomics using RNA sequencing (RNAseq) is the most recent and most
challenging ‘omics sequencing method, but can potentially provide more information
about the adaptations of the microbes to their environment. There are relatively few
studies employing metatranscriptomics due to the difficulty in sample collection to
maintain RNA integrity, and overcoming the overabundance of rRNA molecules
(upwards of 95% of the total RNA content in a bacterial cell) compared to the mRNA
molecules of interest (226). Many of the first studies came from marine and soil
microbiome studies (227-229) and only more recently have studies emerged evaluating
the metatranscriptome of microbial communities associated with human health (230,
231). With the approach only beginning to gain traction, the computational analysis tools
and statistical framework are under active development and multiple methods are used to
analyze transcriptional data.

Evaluating metatranscriptome data has to take in account not only the fluctuating
transcript abundances, but also those of the fluctuations in the bacterial populations. Like
metagenomics, metatranscriptomic reads can be assembled into larger fragments for
annotation (232), but most strategies leverage sequenced genomes to map RNA reads to
annotated genes, or bin sequence fragments by function. Like 16S rRNA data, once reads
are binned (by OTU, gene, or function) the data are represented in a table of read counts
per bin per sample.

Often the goal of RNAseq is to compare two or more conditions or time points in order to
evaluate differential gene expression. There are two important concepts for consideration
when attempting differential RNAseq analysis: the within-condition variation and the
between-condition variation. Sources of variation include technical (sample preparation,
RNA extraction, sequencing library preparation, different sequencing runs, sampling
error, sampling depth/number of reads per sample), and biological (the underlying
differences between the samples or groups of samples). Tools for differential RNAseq often use parametric models to estimate variation. Poisson models have been shown to adequately model variation between technical replicates (233), but the higher biological variation is better accounted for with a dispersion estimate by the negative binomial model (234). Unfortunately for parametric tools, they only work as well as the data fits the underlying distribution. In a comparison of parametric methods PoissonSeq (Poisson model), DESeq, and edgeR (negative binomial models), Li and Tibshirani (235) found that outliers had significant impact on the performance with true false discovery rate (FDR) becoming “unacceptably high”, with the FDR being greatly underestimated. Using RNAseq data from cervical tumours and healthy control tissue, Li and Tibshirani also note many of the top significant features reported by edgeR were unlikely to follow a negative binomial distribution. Several of these features had a large number of reads in one sample of the group, while other samples in the group had few or no read counts. Although these example data come from human tissues, the same situation is arguably even more likely to occur in bacterial metaRNAseq data where, on top of biological disparity in gene expression between samples, variations in the community membership or abundances can affect the underlying genetic potential for the samples being compared. To overcome limitations of parametric approaches, a few non-parametric approaches have been proposed (NOIseq (236), SAMseq (235) which do not make assumptions about the underlying distribution of the count data. However, these methods require large sample sizes to have power (235) – a rare commodity for RNAseq experiments due to the prohibitively high cost of sample prep and sequencing. With the available tools being inadequate for the considerable variation in metaRNAseq analysis, we (myself included) developed a new approach called ALDEx (ANOVA-Like Differential Expression) (237) that explicitly identifies differential features with higher between-condition expression differences (i.e. the difference between non-BV and BV) than the within-condition expression differences (i.e. the differences between samples within the BV group).

Other than good study design, with any of the aforementioned sequencing methods for describing the microbiome the greatest challenges are the computational analyses and
choosing the appropriate tools and statistical framework for the type of data and the hypotheses being addressed.

1.9 Scope and objectives of the thesis

At the outset of my thesis project there were very few studies using high-throughput sequencing methodologies to describe the vaginal microbiome. I started with the general hypothesis that understanding the composition and role of the vaginal microbiome would provide insights into BV, which is a poorly understood condition. I was involved in developing a pipeline using Illumina V6 sequencing in order to describe the vaginal profiles of HIV+ African women with and without BV (48, 197). This methodology was used in Chapter 1 and Chapter 5 to characterize the vaginal microbiota under different conditions. In Chapter 1, the objective was to associate the vaginal profiles of post-menopausal women with the symptoms of vaginal atrophy – a prevalent problem for this cohort. I found that, similar to BV, abundant *Lactobacillus* was associated with healthy asymptomatic women while women who were experiencing symptoms of vaginal dryness had a lower abundance of lactobacilli and higher abundances of other genera (*Gardnerella, Prevotella, Streptococcus, Veillonella*). These dysbiotic profiles are associated with changes in gene expression of the vaginal epithelium (data contributed by Jordan Bisanz). Chapter 5 presents a follow-up to two clinical trials evaluating the effectiveness of probiotic adjunct therapy with standard antibiotic treatment for BV (132) and anti-fungal therapy for VVC (135). The goal was to evaluate the vaginal microbiota to determine if the conjoint treatment and clinical success resulted in changes to the microbiota. This is the first report using high-throughput sequencing to describe the bacterial community in relation to acute VVC. When women recovered from BV, the abundance of *Lactobacillus* was higher and the predominant species was *L. iners*.

A common thread of the microbiota data examined was that *L. iners* is the predominant vaginal organism that can be found in association with healthy and aberrant conditions. In Chapter 3, I used whole genome sequencing on strain AB-1, a vaginal isolate of *L. iners*, to determine if there were elements in its genome that would explain its adaptability to the vagina. The secondary goal was to build a high-quality reference genome for future
studies targeting particular genes or evaluating the transcriptome as presented in Chapter 4.

Based on the apparent plasticity of *L. iners’* genome, and its ability to persist in a variety of conditions, I hypothesized that this organism could be altering its gene expression in response to the environmental changes. Chapter 4 explores this concept by using metaRNAseq on four samples: two women with BV and two who were healthy and asymptomatic. Along with the description of the *L. iners* transcriptome, I sought to address the functional contribution of the whole community. In order to do so, I employed a tool developed by our group (237) that explicitly accounts for within and between condition variance and reports differences that are conserved within a condition. The goal of using this tool was to find functions that are conserved in healthy or BV states, and differentially abundant compared to the other condition. This allowed me to address the hypothesis of whether there are conserved functions in BV despite the differences in the community composition between samples.

Overall, these studies present a comprehensive picture of the vaginal microbiota composition in different conditions. The studies have resulted in several computational tools and pipelines for evaluating 16S and RNAseq data, and for functional annotation by sequence information. The importance of understanding the functional contribution of the organisms is underscored, and resultant data will provide insight into future approaches for modulating the vaginal ecosystem.

### 1.10 References


Chapter 2

2  Vaginal microbiome and epithelial gene expression of post-menopausal women with moderate to severe dryness

This chapter is reproduced with permission (Appendix A) from:


Supporting material for this chapter including Tables and Figures (labeled as S#) are available online:
http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0026602#s5

2.1 Introduction

The onset of menopause is accompanied by a dramatic increase in reported symptoms of vaginal dryness, soreness, irritation or itching, pain with intercourse and bleeding after intercourse (1). Collectively these symptoms affect 25-50% post-menopausal women (2, 3) and significantly impact their quality of life. Vulvovaginal atrophy (VVA) is somewhat of a catchment term for these symptoms and is diagnosed by an assessment of vaginal dryness, irritation, soreness, and dyspareunia with urinary frequency, urgency, incontinence, and the presence of pale and dry vulvovaginal mucosa with petechiae, along with pH >4.6 (4). Though the use of estrogen replacement therapy (ERT) can rapidly reverse some of these changes (5), a large proportion of women are hesitant to use estrogens, even when applied topically (3). Interestingly, a proportion of women appear to remain symptom free, long after the onset of menopause.
A lactobacilli-dominated vaginal microbiota is associated with retention of health, but with menopause it has long been assumed that ERT is required to maintain the dominant lactobacilli and reduce the risk of infection (5). This is supported by studies of the vaginal microbiota using denaturing gradient gel electrophoresis (DGGE) (6, 7); a technique that separates different sequences on a polyacrylamide gel based on the denaturing chemicals and heat applied to the double-stranded PCR products. Next-generation deep sequencing provides a means to uncover the extent of bacterial presence in a given niche, including the vagina (8-10).

The aim of the present study was to identify differences in microbiome composition and epithelial gene expression between post-menopausal women who were clinical asymptomatic compared to women who were suffering vaginal dryness.

### 2.2 Materials and Methods

#### 2.2.1 Study population

The institutional review board of the University of Western Ontario, London, Canada, approved the protocol. The study was registered at clinicaltrials.gov with ID# NCT01084616.

From a practice with over 500 post-menopausal women, a total of 32 patients between ages 42 and 77 were identified and recruited at the Victoria Family Medical Center in London, Canada between January and February 2010, and followed until June 2010. Participants were recruited to the study via their physician (Jo-Anne Hammond), or information sheets posted in the doctors’ offices. All participants had at least 12 months of amenorrhea, had not used antibiotics or had any vaginal infection one month prior to screening, had not used systemic estrogen products 6 months prior to sampling, or a topical estrogen-containing product one week before sampling. Women were excluded if they had used douches or lubricants in the past week.

After giving informed consent, each participant completed a questionnaire on demographics, history of vaginal or urinary tract symptoms or infections, and a self-assessment of the severity of five vaginal symptoms: vaginal dryness, vaginal irritation or
itching, pain during urination not associated with infection, vaginal soreness, pain during sexual intercourse, bleeding after sexual intercourse. Symptoms were rated as none, mild, moderate, or severe. Pain with intercourse and bleeding after intercourse could be scored as “not applicable”.

A gynecologic examination was performed by a nurse trained to detect dryness and atrophy to provide an “observed” assessment of vaginal color, dryness, blanching and integrity of tissue, vaginal tissue petechiae, and overall vaginal atrophy. Vaginal pH was measured using the pHem-Alert indicator (Gynex, Wyoming, USA). A cytobrush (Cytobrush plus GT, Cooper Surgical Inc. USA) was used to collect epithelial cells from the mid-vagina for human mRNA analysis. The cytobrush was immediately placed in a 1.5 ml sterile RNase-free microtube containing 700 µl RNA later stabilization agent (Applied Biosystems, Austin, Texas, USA) and stored at 4°C within 4 hours of collection. RNA extraction occurred within 72 hours of collection. Two polyester Dacron swabs were used to sample the mid-vagina for microbial analysis: one was archived at -80°C for later bacterial DNA extraction, and the other was rolled on a microscope glass slide and scored for bacterial vaginosis (BV) using the Nugent criteria (11). Sixteen participants with moderate to severe vaginal dryness were followed every 2 weeks for up to 10 weeks total.

2.2.2 Bacterial DNA extraction, amplification, and sequencing

Bacterial DNA was extracted using Instagene (Bio-Rad) and the V6 region of the 16S rRNA gene was PCR amplified as described previously (9, 12) using barcoded primers: L-V6 (5’-CAACGCGARGAACCTTACC-3’) and R-V6 (5’-ACAACACGAGCTGACGAC-3’). Based on the intensity of the ethidium bromide-stained band on an agarose gel (estimated by eye), an approximate equimolar ratio of PCR products were mixed together to give the final sample sent for Illumina paired-end sequencing at The Centre for Applied Genomics in Toronto, Canada.

2.2.3 OTU clustering and taxonomic assignment

Sequence filtering, processing, and microbial analysis was performed as described previously (9, 12) and summarized as follows. After overlapping paired reads to get full-
length V6 sequence, the reads were clustered at 95% nt identity using Uclust version 3.0.617 (http://www.drive5.com/usearch/index.html). The most abundant sequence in a cluster was selected as a representative OTU. Taxonomic assignments for the representative OTUs were made through alignments against the Greengenes (13) database using NAST (14) and further manual curation by BLAST to the NCBI non-redundant database. The representative OTU sequences occurring with at least 1% abundance in any one sample have been deposited in GenBank under accession numbers JF262791-JF262909.

2.2.4 Vaginal epithelial RNA extraction

Each of the 32 subjects provided a single sample for human genomic array. Vaginal cytobrush samples were centrifuged (5000 xg, 10min, 4°C) and the supernatant was discarded before RNA extraction by TRIZol following the manufacturer’s protocol (Invitrogen, Carlsbad, CA, USA). The pellet was resuspended in 20 µl RNase-free water and stored at -80°C until further analysis. RNA quantity and quality was assessed by a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) to ensure a 260nm/280nm ratio above 1.8 and a 260nm/230nm ratio above 1.4 before quality was confirmed using an Agilent 2100 Bioanalyzer (Agilent technologies, Santa Clara, CA, USA). All steps of GeneChip probe preparation and hybridization were performed by the LRGC using standard protocols. All liquid handling steps were performed by a GeneChip Fluidics Station 450 and GeneChips were scanned with the GeneChip Scanner 3000 7G (Affymetrix, Santa Clara, CA) using Command Console v1.1.

Sufficient quality (RIN ≥ 7) and quantity of RNA was received from 10 of the samples, four from the dryness group (age 57.0 ± 8.9 years, range: 50-70 years; with an average 13.3 ± 11.4 years since last menses, range: 3-21 years) and six from the subjects with a normal Nugent score (age 56.2 ± 9.3 years, range: 48-74 years; with an average time since last menses of 11.3 ± 9.8 years, range: 6-30 years).
2.2.5 Gene expression analysis

Probe level (.CEL file) data was generated using Affymetrix Command Console v1.1. Probes were summarized to gene level data in Partek Genomics Suite v6.5 (Partek, St. Louis, MO) using the RMA algorithm (15).

Gene expression data has been deposited in the Gene Expression Omnibus Database under accession number GSE26761.

2.2.6 Statistical analyses

Shannon’s diversity index (16) was calculated for each subject at each time point as the abundance values of 119 operational taxonomic units (OTUs). The Shannon’s diversity index \( S \) was defined as 

\[
S = - \sum_{i=1}^{119} p_i \log(p_i)
\]

where \( p_i \) denotes the proportional abundance values at the \( i \)th OTU in each sample. The \textit{Lactobacillus} ratio was calculated for each subject at each time point, as the ratio of total abundance value for the \textit{Lactobacillus} OTUs versus all the OTUs.

Hierarchical clustering was performed by calculating the Euclidean distance between microbiota samples using the hclust function in R (17) and the similarity is plotted in a dendrogram in Figure 2-1.
Figure 2-1. Microbiota profiles for 32 post-menopausal women clustered by biota similarity

Each bar represents a single vaginal sample, and the colored segments represent the relative fraction of each bacterial taxon detected at 1% relative abundance or greater in any one sample. Sequences at less than 1% abundance have been included in the “remainder” fraction at the top of the bar (see color legend of bacterial taxa). The microbiota are clustered by similarity as represented in the dendrogram above. The sample name (participant ID-time point) is labeled in the dendrogram and corresponds to the bar below. The dryness score as observed by the examining nurse is represented below each microbiota bar. Color taxa legend can be found in Figure 2-2.
Figure 2-2. Color taxa legend for Figure 2-1 and Figure 2-3

Starting with the left column from bottom up, the order of the colors matches the order of the barplots read from bottom to top.

To determine if the linear trends in the $S$ trajectories differ significantly across subjects, the following linear mixed effect model (18) was applied to the data $Y_{it} = \mu + a_i + b_i t + \epsilon_{it}$. The fixed effect being the intercept only and random effects being subject intercepts and slopes where $Y_{it}$ denoted the $S$ value of subject $i$ at time point $t$ and $\mu$ is an unknown fixed constant, $a_i$ is a random subject intercept, $b_i$ a random subject slope, and $a_i$, $b_i$, and $\epsilon_{it}$ are independent normal random variables.

In order to determine if a statistical relationship existed between microbial diversity and vaginal dryness, a weighted least squares analysis (19) was applied to the data. Observed dryness scores of “none”, “mild”, “moderate”, and “severe” were treated as ordinal values 0, 1, 2, and 3, respectively. The weighted least squares model $\bar{Y}_{id} = \mu_0 + \mu_1 d + \epsilon_{id}$ was used to test the linear effect of observed dryness with respect to microbial diversity. Where $\bar{Y}_{id}$ denotes the average $S$ of subject $i$ at dryness $d$. The quantities $\mu_0$ and $\mu_1 d$ are constants representing intercept and slope respectively. We also applied the Mann-Whitney-Wilcoxon rank-sum to the OTUs by comparing the relative abundances between women with and without dryness at time point 0.
2.3 Results

The reported and observed vaginal symptoms are presented in Figure S1. The observed vaginal dryness correlated well with observed atrophy \( (r^2 = 0.84) \) supporting vaginal dryness as an appropriate proxy for the occurrence of vaginal atrophy. Moderate to severe scores were in agreement for three out of the four measures that included dryness, discoloration, blanching and petechiae in 11 participants.

Illumina sequencing resulted in 1,758,430 total V6 reads after quality filtering, with an average of 17410 ± 9917 reads per sample (range: 2,123-53,054). Figure 2-1 shows the microbiota findings for the 32 participants at the first sample time point. Using a cutoff of 1% abundance in any sample, there were 119 distinct operational taxonomic units (OTUs) identified, each representing a bacterial taxon.

Sixteen of the 32 participants were selected at their initial visit and followed for microbiota sampling every 2 weeks for 10 weeks total. This time series showed that the microbiome abundance profiles did not fluctuate greatly within a subject (Figure 2-3). When a linear mixed effect model was applied to the Shannon’s diversity index data, a statistical linear trend was not observed \( (p = 0.364) \) indicating that the abundance of microbial diversity from week to week did not change significantly within a subject. The biotas of women who were least symptomatic (“none” or “mild” dryness score) had low bacterial diversity with a dominance of lactobacilli (Figure 2-1 and Figure 2-3). Conversely, women with “moderate” or “severe” dryness have a decreased abundance of lactobacilli and a large diversity with number of species detected. In this group of women experiencing vaginal dryness, there was a high representation of *Prevotella timonensis* (OTU_6), *Porphyromonas* (OTU_9), *Peptoniphilus* (OTU_27), and *Bacillus* (OTU_34) using a Mann-Whitney-Wilcoxon rank-sum test (Supporting Figure S2). A weighted least squares model was applied to the data and a statistical linear trend was observed \( (p = 0.00141) \) indicating an inverse correlation between *Lactobacillus* ratio and dryness.
Figure 2-3. 16S (V6) microbiota profiles for 16 post-menopausal women sampled every 2 weeks

Each bar represents a single vaginal sample, and each cluster of bars is a single participant (starting at time 0 and sampled every 2 weeks for up to 10 weeks total). The colored segments represent the relative fraction of each bacterial taxon detected at 1% relative abundance or greater. Sequences at less than 1% abundance have been included in the “remainder” fraction at the top of the bar (see color legend of bacterial taxa). The dryness score as observed by the examining nurse is represented below each microbiota bar. Sample time points that were included in the microarray analysis are marked with an arrowhead: the first six green arrows are controls (no or mild dryness), and the last four red arrows are women experiencing moderate to severe dryness. Color taxa legend can be found in Figure 2-2.

Lactobacilli were dominant in many women, irrespective of age (participant 13 age 74 versus participant 6 age 42), and in absence of ERT. We noted lactobacilli at an abundance of $\geq 10\%$ in 29/32 (91%) women at time zero, while 17/32 (53%) had lactobacilli as the overall dominating organism (greater than 50% relative abundance).

In order to perform human genomic array analysis, one sample was collected at random from each subject. Based upon successful high-quality RNA isolation, 10 resultant samples were divided into healthy controls and subjects experiencing vaginal dryness based on the observed dryness score (Figure 2-3).

After importing intensity data into Partek GS 6.5, an ANOVA was applied to identify differentially expressed genes between the control and dryness groups. The list of genes
was filtered to only include those differentially expressed by at least 2-fold ($p < 0.05$), which yielded a list of 960 probe sets used for subsequent comparative analyses (Table S1). A heatmap of the filtered list (Figure 2-4) shows two main branches separating the dryness and control groups. Participant 4 had an intermediate gene expression profile between the dryness and control groups despite having no self-reported or observed signs of dryness and correspondingly no additional signs of atrophy.

**Figure 2-4. Heatmap of vaginal epithelial gene expression of 10 samples**

A clustered heatmap of differential gene expression (>2-fold change, $p<0.05$) between the control and dryness groups. Samples are labeled as participant number-time point (week). Samples assigned to the dryness or control groups based on physiological examination of the vagina are well separated by gene expression differences with exception of 4-8 which has an intermediate gene expression profile and clusters closer to the dryness group.

Then, the top 20 genes in either direction with the greatest fold change between the dryness and control groups were further analyzed (Table 2-1) to determine where the
most dramatic changes occurred. The most up-regulated genes in the dryness group were MMP7 (13.75-fold), and the choline transporter SLC44A4 which was interrogated by three separate probe sets, all up-regulated 9.19-fold with identical \( p \)-values. The most down-regulated genes in the dryness group were SPINK7 (-38.65-fold) and TGM3 (-21.86-fold).

### Table 2-1. The top 20 genes up- and down- regulated genes (\( p<0.05 \)) in the vaginal dryness group compared to controls.

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Assignment</th>
<th>RefSeq ID</th>
<th>( p )-value</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMP7</td>
<td>matrix metallopeptidase 7 (matrilysin, uterine)</td>
<td>NM_002423</td>
<td>7.73E-03</td>
<td>13.75</td>
</tr>
<tr>
<td>SLC44A4</td>
<td>solute carrier family 44, member 4</td>
<td>NM_025257</td>
<td>2.23E-02</td>
<td>9.19</td>
</tr>
<tr>
<td>SLC44A4</td>
<td>solute carrier family 44, member 4</td>
<td>NM_025257</td>
<td>2.23E-02</td>
<td>9.19</td>
</tr>
<tr>
<td>SLC44A4</td>
<td>solute carrier family 44, member 4</td>
<td>NM_025257</td>
<td>2.23E-02</td>
<td>9.19</td>
</tr>
<tr>
<td>CFH</td>
<td>complement factor H</td>
<td>NM_000186</td>
<td>4.79E-03</td>
<td>8.72</td>
</tr>
<tr>
<td>PIGR</td>
<td>polymeric immunoglobulin receptor</td>
<td>NM_002644</td>
<td>2.01E-02</td>
<td>8.64</td>
</tr>
<tr>
<td>IL19</td>
<td>interleukin 19</td>
<td>NM_153758</td>
<td>9.50E-03</td>
<td>8.31</td>
</tr>
<tr>
<td>CFH</td>
<td>complement factor B</td>
<td>NM_001710</td>
<td>4.69E-02</td>
<td>7.33</td>
</tr>
<tr>
<td>PLAT</td>
<td>plasminogen activator, tissue</td>
<td>NM_000930</td>
<td>2.23E-02</td>
<td>7.19</td>
</tr>
<tr>
<td>CFB</td>
<td>complement factor B</td>
<td>NM_001710</td>
<td>2.08E-03</td>
<td>7.10</td>
</tr>
<tr>
<td>WFDC2</td>
<td>WAP four-disulfide core domain 2</td>
<td>NM_006103</td>
<td>8.32E-03</td>
<td>6.93</td>
</tr>
<tr>
<td>CXCL6</td>
<td>chemokine (C-X-C motif) ligand 6</td>
<td>NM_002993</td>
<td>1.81E-02</td>
<td>6.50</td>
</tr>
<tr>
<td>TSPAN1</td>
<td>tetraspanin 1</td>
<td>NM_005727</td>
<td>2.08E-04</td>
<td>6.30</td>
</tr>
<tr>
<td>AGR2</td>
<td>anterior gradient homolog 2 (Xenopus laevis)</td>
<td>NM_006408</td>
<td>1.50E-02</td>
<td>6.18</td>
</tr>
<tr>
<td>TRIM31</td>
<td>tripartite motif-containing 31</td>
<td>NM_007028</td>
<td>4.77E-02</td>
<td>5.85</td>
</tr>
<tr>
<td>ASS1</td>
<td>argininosuccinate synthetase 1</td>
<td>NM_000050</td>
<td>3.05E-02</td>
<td>5.70</td>
</tr>
<tr>
<td>PLAC8</td>
<td>placenta-specific 8</td>
<td>NM_016619</td>
<td>2.03E-02</td>
<td>5.51</td>
</tr>
<tr>
<td>SPINK7</td>
<td>serine peptidase inhibitor, Kazal type 7 (putative)</td>
<td>NM_032566</td>
<td>4.81E-04</td>
<td>-38.65</td>
</tr>
<tr>
<td>TGM3</td>
<td>transglutaminase 3</td>
<td>NM_003245</td>
<td>4.24E-05</td>
<td>-21.86</td>
</tr>
<tr>
<td>SBSN</td>
<td>Suprabasin, part of a gene complex including dermokine</td>
<td>NM_198538</td>
<td>3.28E-05</td>
<td>-20.93</td>
</tr>
<tr>
<td>ALOX12</td>
<td>arachidonate 12-lipoxygenase</td>
<td>NM_000697</td>
<td>4.43E-04</td>
<td>-19.61</td>
</tr>
<tr>
<td>KPRP</td>
<td>keratinocyte proline-rich protein</td>
<td>NM_0010252</td>
<td>2.30E-04</td>
<td>-17.43</td>
</tr>
<tr>
<td>GYS2</td>
<td>glycogen synthase 2 (liver)</td>
<td>NM_021957</td>
<td>1.66E-03</td>
<td>-16.78</td>
</tr>
<tr>
<td>DSG1</td>
<td>desmoglein 1</td>
<td>NM_001942</td>
<td>5.02E-03</td>
<td>-16.61</td>
</tr>
</tbody>
</table>
To assess the relationship of differentially expressed genes, a GeneOntology (GO) enrichment analysis of the filtered list of 960 probe sets was performed using Partek GS 6.5. The top 10 GO-functions enriched are displayed in Table 2-2, and demonstrate changes in epithelial remodeling and immune response. Cornified envelope had the highest enrichment score with 62.50% (10 genes) in this process and was down-regulated in the dryness group.

Table 2-2. Enrichment of GO terms in the filtered gene list (n=960).

<table>
<thead>
<tr>
<th>Function</th>
<th>GO ID</th>
<th>Enrichment Score</th>
<th>Enrichment p-value</th>
<th>No. of Genes</th>
<th>% of total genes in GO term group differentially expressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cornified envelope</td>
<td>1533</td>
<td>21.81</td>
<td>3.37E-10</td>
<td>10</td>
<td>62.50</td>
</tr>
<tr>
<td>Keratinocyte differentiation</td>
<td>30216</td>
<td>20.01</td>
<td>2.04E-09</td>
<td>11</td>
<td>47.83</td>
</tr>
<tr>
<td>Immune response</td>
<td>6955</td>
<td>19.20</td>
<td>4.60E-09</td>
<td>38</td>
<td>13.52</td>
</tr>
<tr>
<td>Cell differentiation</td>
<td>30154</td>
<td>17.96</td>
<td>1.59E-08</td>
<td>24</td>
<td>17.91</td>
</tr>
<tr>
<td>Epidermis development</td>
<td>8544</td>
<td>17.55</td>
<td>2.38E-08</td>
<td>17</td>
<td>23.94</td>
</tr>
<tr>
<td>Tissue development</td>
<td>9888</td>
<td>17.48</td>
<td>2.55E-08</td>
<td>23</td>
<td>18.11</td>
</tr>
<tr>
<td>Cellular developmental process</td>
<td>48869</td>
<td>15.29</td>
<td>2.28E-07</td>
<td>30</td>
<td>13.45</td>
</tr>
</tbody>
</table>
### Table: Gene expression values

<table>
<thead>
<tr>
<th>Region</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracellular region</td>
<td>5576</td>
<td>13.99</td>
<td>8.37E-07</td>
<td>37</td>
<td>11.28</td>
</tr>
<tr>
<td>Epithelial cell differentiation</td>
<td>30855</td>
<td>13.12</td>
<td>2.01E-06</td>
<td>6</td>
<td>60.00</td>
</tr>
<tr>
<td>Innate immune response</td>
<td>45087</td>
<td>13.03</td>
<td>2.18E-06</td>
<td>11</td>
<td>26.83</td>
</tr>
</tbody>
</table>

#### 2.4 Discussion and Concluding Remarks

The microbiota of this set of post-menopausal women was quite stable with 1 out of a possible 62 transitions between normal and BV Nugent scores, unlike another study of pre-menopausal women that observed 226 transitions out of a possible 1365 opportunities ($p = 0.003$) (20). If indeed further studies confirm that the microbial profiles of post-menopausal women are relatively stable, perhaps it is due to lack of menstruation which is known to cause stage-dependent variations in the microbiota (21).

The Illumina sequencing method identified double or triple the number of bacterial types per subject compared to previous denaturing gel electrophoresis methods (7). Similar to recent deep sequencing studies of pre-menopausal women (8-10, 22, 23), we noted *Lactobacillus iners* and *Gardnerella vaginalis* were universally present and possibly represent core members of the vagina. However, there were much lower abundances (<1%) of *Mobiluncus, Staphylococcus, Sneathia, Bifidobacterium*, and *Gemella* in our study set of post-menopausal women. This could be due to our smaller sample size rather than micro-environmental changes associated with menopause and so the physiological significance is not known. As more information emerges about the different organisms in the microbiota, the types of clones and virulence properties of them, for example differences in the expression of various toxins and mucus-degrading enzymes by *G. vaginalis* (24), will allow a better understanding of which strains equate to health versus disease.

Vaginal dryness, a condition that is associated with significant changes in the transcription of genes associated with cellular structure and immune function as shown here by the human microarray data, is a symptom of atrophy which afflicts large numbers of menopausal women and has been linked with depletion of estrogen (25). The muscular
layer of the vaginal wall tends to be thicker in postmenopausal than premenopausal women, while the epithelial layer is much thinner (26). Having developed a method to use Affymetrix arrays to read epithelial cell gene expression changes in the vagina (27), we applied this here and found down-regulation of genes involved in maintaining proliferation and barrier function of the vaginal epithelium in the dryness group. Without a cornified envelope, epithelial cells are much more sensitive to drying, chemical agents and microbial infection (28). Many of the genes in this group encode structural components of the cornified envelope so lowered expression in the dryness group implies reduced structural integrity of the tissue. This is mirrored in the down-regulation of adhesion molecules such as SBSN and DSG1, implicated in maintaining proper epithelial barrier function (29, 30). Similarly, ALOX 12 (-19.61-fold) has also been shown to be a positive regulator of epidermal barrier function (31), and keratin is an important structural molecule in the epidermis imparting mechanical stability to the tissue (32).

The most down-regulated gene in the dryness group was SPINK7 (alias ECRG2), a member of the SPINK family of proteases, which is known to regulate invasion and migration by preventing extracellular matrix (ECM) degradation through inhibiting the action of urokinase-type plasminogen activator (33). Matrilysin-1 protein (MMP) 7 opposes the SPINK family proteins, and its up-regulation by 13.75-fold and cleaves ECM components and up-regulate inflammation (34). Thus, it is possible that the combination of increased ECM degradation with decreased inhibition of protease inhibitors contributes to vaginal thinning.

Both complement factors B and H (CFB and CFH) and other components of the complement system were up-regulated in the dryness group. Since dryness is accompanied by inflammation, this suggests that the up-regulation of the regulatory molecules (CFH and CFI), although perhaps capable of protecting host cells, is not sufficient to dampen the inflammatory response. This potential for dysregulation of the complement system in vaginal dryness warrants further study.

Differential expression of many cytokines and chemokines, the messengers of inflammation and chemotaxis, was noted. Three CXCL-family chemokines, known to
recruit monocytes and leukocytes to the vaginal epithelium, were observed to be up-regulated 4.11 to 6.5 fold in the dryness group. Polymeric immunoglobulin receptor (PIGR) was up-regulated 8.64-fold in the dryness group; indicative of increased mucosal immune activity.

Two highly up-regulated genes, SLC44A4 and PLAT, could be associated with the symptoms of vaginal dryness and atrophy. The SLC44A4 gene product is a sodium-dependent transmembrane transport protein involved in the uptake of choline in neurons, which is used in neurons in the production of acetylcholine, which is one of the main neurotransmitters of the parasympathic nervous system (35). As the vaginal epithelium is rich in nerve endings and these can increase in density with reduced estrogen (35), it is possible that increased activity of this transporter may contribute to the sensation of pain and irritation through cholinergic nerve endings. The PLAT gene product is a serine protease which converts plasminogen to plasmin, and functions in cell migration and tissue remodeling as well as reducing blood clotting. Up-regulation of this gene may play a role in the increased bleeding and tissue sensitivity associated with vaginal dryness.

Our gene array findings agree with data obtained by Cotreau et al. (36) on vaginal biopsies from 19 women with vulvovaginal atrophy (VVA) before and after treatment with 17\(\beta\)-estradiol. The similarity in results suggests that many of the genes being differentially regulated between dryness and healthy women may be as a result of differential estrogen levels between the groups. This is further supported by the up-regulation of one of the vaginal epithelial cornified envelope genes (LCE3D) following estrogen replacement therapy (37), that was 14-fold down-regulated in women with vaginal dryness.

In summary, the vaginal microbiota of 16 post-menopausal women showed little fluctuation over time, unlike that of pre-menopausal women reported previously. There was an inverse correlation between *Lactobacillus* ratio and dryness, a condition commonly found after menopause, which shown here to be associated with changes in vaginal epithelial cell integrity and inflammation.
2.5 References


17. R Development Core Team (2013) R: A Language and Environment for Statistical Computing


Chapter 3

3 At the crossroads of vaginal health and disease: the genome sequence of *Lactobacillus iners* AB-1

This chapter is reproduced with permission (Appendix B) from:


Supporting material for this chapter including Tables and Figures (labeled as S#) are available as a PDF online:


3.1 Introduction

The vaginal microbiota of premenopausal women with no symptoms or signs of disease comprises primarily *Lactobacillus* species. The consistency of this finding is quite remarkable given the several hundred other microbial species that have been detected in the vagina (1-5) and the accessibility of even more from the rectal-anal skin and through bathing and sexual contact. For decades, lactobacilli have been regarded as beneficial to the vagina by preventing infections through production of organic acids, hydrogen peroxide (H2O2) and other anti-microbial substances, and by the fact that depletion of *Lactobacillus* populations often coincides with bacterial disease onset.

The fastidious *Lactobacillus iners* is consistently the most common *Lactobacillus* sp. in the vagina (1, 3, 6, 7), yet little is known about its characteristics. The detection of *L. iners* in women with, and recovering from, bacterial vaginosis (BV) has led to the suggestion that it is not protective against disease (8). However, unlike other species that appear to be easily displaced by pathogens and infectious conditions, *L. iners’* ability to
Persist (9) may prove it to be important commensal, which may have a role in recovery of the microbiota, post-disease resolution. Interestingly, through 16S microbial community profiling, the L. iners sequences recovered from different women show homogeneity and suggests a lack of strain diversity in this species (10). Here we report the genome sequence and annotation of a representative of the species, L. iners AB-1, and through comparative genomics, show that it is the smallest lactobacilli discovered to date (11) differing markedly from intestinal and other urogenital species.

### 3.2 Materials and Methods

#### 3.2.1 Organism

*Lactobacillus iners* AB-1 is a vaginal isolate from a healthy woman. For DNA isolation for genome sequencing, it was cultured on Columbia blood agar (VWR, DF0792-17) supplemented with 50ml/L sheep’s blood (Cedarlane, DSB050) at 37°C under strict anaerobic conditions.

#### 3.2.2 Genome assembly

We received 621,824 reads from 454 pyrosequencing ranging between 100bp and greater than 600bp (Next-Gen Sequencing Facility at Iowa State University, Iowa). The reads were trimmed for low quality bases and leftover adapter sequences using the python script, sff_extract.py (downloadable at http://bioinf.comav.upv.es/sff_extract/index.html). Of the total reads after trimming, the 68,626 reads greater than 500bp were used to generate a preliminary assembly using the Minimus assembler (AMOS package) (12). The resulting contigs were used as long reads in combination with the approximately 12 million short (35nt) paired reads from Illumina (Illumina, Inc. San Diego, California) for paired-end assembly by the Velvet assembly program (13). Contigs greater than 200bp were retained, and these 47 contigs were used as the backbone for gap closure and scaffolding.

#### 3.2.3 Gap closure and scaffolding

To close gaps between contigs, a method of “read walking” was employed. A BLAST database of all 454 reads greater than 500bp was created using formatdb. Contig ends
were BLASTed (14) against the database to find reads with a perfect overlap of at least 300bp. If no reads matched, the end of the contig was trimmed back until a match could be found, or the required overlap was reduced. Reads matching the ends of contigs were compared to the database of 454 reads using BLAST in order to extend the contig ends until another contig was reached. If no matching reads were uncovered in the dataset of reads greater than 500bp, a second database including reads greater than 400bp (262,847 reads) was used. Once inter-contig reads were obtained, gaps between contigs were joined using Minimus (12). This read-walking approach was similarly used to place the rRNA operons by walking out of the contigs identified as 16S and 23s sequence by BLAST to the NCBI non-redundant database. Reads matching the ends of these contigs were imported into JalView (15) where the multiple sequence alignments could be partitioned into separate sequence paths into contig ends. Additional contig joinings used sequence information from an unfinished Lactobacillus iners project in the GenBank database (GenBank accession number: ACLN00000000). Mauve was used to evaluate the alignment and contig order between the two datasets (16). Initially, a coverage of 5 reads was required to extend a contig, and upon completion of gap filling all contigs were evaluated for expected sequence coverage and “mate-happiness” between Illumina reads as described below.

By read-walking, the genome was assembled into seven contiguous pieces which were then ordered into a single scaffold. Scaffolding order was determined based on maximum synteny with the most closely related organisms, L. johnsonii and L. gasseri, using the PGA platform (17). The contigs were also oriented and ordered based on expected positions of the rRNA operons (18). Optical Mapping (OpGen, Inc., Madison, Wisconsin) validated contig order and orientation, as well as the expected genome size of 1.3Mb (Figure S6). Primers were designed for the ends of each scaffolded contig and long-range PCR (Expand Long Template PCR System, Roche) was used to verify connections between neighboring contigs as predicted by optical mapping. PCR products were electrophoresed on a 1% agarose gel and bands were isolated, gel purified, and submitted for dideoxy chain termination sequencing (London Regional Genomics Institute, London, Ontario). Sanger reads were used to extend contig ends, and read-
walking was again performed to close remaining gaps. The majority of contig ends were rRNA operons and the PCR and Sanger sequences confirmed these rRNA placements.

### 3.2.4 Validating genome coverage

The final scaffold was independently validated by separately mapping the original Illumina paired reads and the 454 reads and then assessing the expected read coverage. Using Novopaired (www.novocraft.com), 97.7% of the Illumina paired reads mapped back to the scaffold with unique alignment. The expected distance between read pairs, termed “mate-happiness”, was also evaluated from the Novopaired output to identify areas of compression (mates too close together) or expansion (mates too far apart). We used in-house developed software and the BLAST algorithm to map the 454 reads to the scaffold using cutoffs of 98% ID, 95% read length excluding reads with insertions or deletions compared to the scaffold. Coverage and mate happiness for each final scaffold was visualized using R (19) (Figures S8 and S9). One region identified by optical mapping suggested ~12,000 bases missing in the final assembly (Figure 6). This region had good coverage and mate-happiness by the Illumina reads, but low coverage of 454 reads suggesting short tandem repeats which could not be resolved.

### 3.2.5 Annotation

Open reading frames (ORFs) greater than 100nt were predicted using GeneMark (20) and by Glimmer (21). The translated ORF predictions were compared to the NCBI non-redundant database (nrdb) using BLASTp to evaluate gene predictions. There were several cases of GeneMark predicted ORFs that were shorter in length compared to homologous sequences in the database, and the corresponding Glimmer prediction better matched the length of the proteins in the database. In these instances, the Glimmer prediction was preferentially retained over the GeneMark prediction in order to prevent overestimation of truncated pseudogenes.

Several tools combined with custom-created Perl scripts were used for the manual annotation of predicted ORFs including: BLAST to NCBI nrdb, COG (Clusters of Orthologous Groups of proteins) (22), and LaCOG (Lactobacillales-specific Clusters of Orthologous protein coding Genes) (23) databases. In addition, automatic annotation
results were collected from RAST (Rapid Annotation using Subsystem Technology) (24). Metabolic predictions were made by KAAS (KEGG Automatic Annotation Server) (25). HMMER (http://hmmer.org/) was used to search the sortase database (http://bamics3.cmbi.kun.nl/sortase_substrates) for sortase substrates (26). The predicted ORFs were also submitted to Pfam (27) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) for conserved domain and transmembrane domain predictions respectively. A list of Pfam domains related to adhesion adapted from Kankainen et al. (28) was used to search the predicted ORFs with $E$ values of $1e^{-3}$ or less considered significant (Table S6). A summary of Surface and exported proteins were identified using SLEP (29). Transporter genes were predicted and annotated by TransAAP (http://www.membranetransport.org/). Transfer RNA (tRNA) genes were identified using tRNAscan-SE (30) Annotation predictions were manually validated.

3.2.6 Prediction of highly expressed genes (using the codon adaptation index, CAI)

Highly expressed genes show a biased codon usage pattern, and the CAI provides an indication of expression (31). Using the EMBOSS (32) cai tool, the ORF sequence of the generally highly expressed small and large subunits of the ribosomal proteins were used as a reference for calculating the codon adaptation index (CAI) of all the predicted genes in the *L. iners* genome.

3.2.7 Comparative genomics

Comparative genomics of *L. iners* was performed largely using sequence acquired from the GenBank database for *L. crispatus* JV-V01 (GenBank accession: ACKR00000000), *L. johnsonii* NCC 533 (NC_005362), *L. gasseri* ATCC 33323 (NC_008530), *L. acidophilus* NCFM (NC_006814), and *L. delbrueckii* subsp. *bulgaricus* ATCC 11842 (NC_008054).

Ortholog predictions for each pair of organisms was predicted by InParanoid (33) using the default settings, and custom Perl scripts were developed in-house to merge the datasets into a four-way comparison of overlapping ortholog predictions. The lists were submitted to the 4-way Venn Diagram Generator.
(http://www.pangloss.com/seidel/Protocols/venn4.cgi) and the predicted values were used to generate a figure. Unique proteins were defined as those without ortholog predictions. For simplicity, excluded from the total count were predicted orthologous groups between organism pairs that were not a one-to-one orthologous pairing (if a protein from organism A had more than one predicted ortholog from organism B, and vice versa). Other homology predictions are based on BLAST results to the NCBI nrdb.

### 3.2.8 Prediction of horizontally acquired genes

Predicted protein sequences from *L. iners* AB-1 and from the organisms listed above were compared to the NCBI nrdb by BLAST. Excluding self-hits and hits to the same species, genes were identified as foreign if the three most significant hits ($E <= 1E^{-20}$) were a genus other than *Lactobacillus* with the most significant hit having at least 60% protein identity to the query sequence.

### 3.2.9 Construction of phylogenetic tree of select lactobacilli

Ribosomal protein sequences were chosen and tested for vertical transmission as outlined by Ciccarelli et al. (34). The following sequences contained clear orthologs in all the organisms and were vertically transmitted: ribosomal protein L1, L10, L11, L13, L15, L16, L17, L18, L2, L20, L21, L22, L23, L24, L27, L29, L3, L4, L5, L6, S10, S11, S12, S13, S20, S3, S4, S5, S7, S8, S9. These concatenated sequences were sent for phylogenetic analysis using the 'one click' mode at www.phylogeny.fr (35). This is a "default" mode which uses a pipeline already set up to run and connect programs recognized for their accuracy and speed (MUSCLE for multiple alignment, Gblocks for alignment curation, PhyML for phylogeny and finally TreeDyn for tree drawing) to reconstruct a robust phylogenetic tree from a set of sequences. The trees of cholesterol-dependent cytolsins (Figures S2 and S3) were created by the same method.

### 3.2.10 Cell-wall protein isolation

Protein isolation was performed using the method adapted from Cole, et al. (36). Briefly, *L. iners* was grown in 25ml MRS broth for 96 hours. The culture was spun down, the supernatant was removed, and the pellet was frozen at -20°C until isolation. After
thawing, the pellet was washed two times using TES buffer, resuspended in a 10% lysozyme solution in TES, and incubated at 37°C with shaking for 2h. After centrifugation, the supernatant containing the solubilized cell wall-associated proteins was collected. Proteins were precipitated using 10% TCA before separating by size by electrophoresis on SDS-PAGE, and bands were visualization by coomassie blue staining. For Western Blot analysis, the SDS-PAGE electrophoresed proteins were transferred to a polyvinylidene difluoride (PVDF) membrane for immuno-blotting as described below.

### 3.2.11 Mass spectrometry

After isolation of proteins, separation of by SDS-PAGE, and coomassie staining as described above, the visibly stained bands were isolated for mass spectrometry. In-gel digestion was performed using a MassPREP automated digester station (PerkinElmer). Gel pieces were Coomassie destained using 50mM ammonium bicarbonate and 50% acetonitrile (silver de-stained using a 50 mM sodium thiosulphate 5 hydrate and 15 mM potassium ferricyanide solution), which was followed by protein reduction using 10 mM dithiothreitol (DTT), alkylation using 55 mM iodoacetamide (IAA), and tryptic digestion. Peptides were extracted using a solution of 1% formic acid and 2% acetonitrile and lyophilized. Prior to mass spectrometry analysis, dried peptide samples were re-dissolved in a 10% acetonitrile and 0.1 % TFA (trifluoroacetic acid).

MALDI matrix, α-cyano-4-hydroxycinnamic acid (CHCA), was prepared as 5 mg/mL in 6mM ammonium phosphate monobasic, 50% acetonitrile, 0.1 % trifluoroacetic acid and mixed with the sample at 1:1 ratio (v/v).

Mass Spectrometry data were obtained using a 4700 Proteomics Analyzer, MALDI TOF TOF (Applied Biosystems, Foster City, CA, USA). Data acquisition and data processing were respectively done using 4000 Series Explorer and Data Explorer (both from Applied Biosystems). The instrument is equipped with a 355 nm Nd:YAG laser; the laser rate is 200 Hz. Reflectron positive ion mode was used and the instrument was calibrated at 50 ppm mass tolerance. Each mass spectrum was collected as a sum of 1000 shots.
Protein identification was made using a local MASCOT (www.matrixscience.com) server and a custom-build peptide mass database constructed from the predicted protein sequences of *L. iners* AB-1. Search criteria allowed for up to 3 missed cleavage sites and a 50 ppm tolerance, and searches were performed using constant carbamidomethyl and variable oxidation modifications.

### 3.2.12 Rabbit polyclonal antisera

Polyclonal antisera was raised in rabbits using formalin-fixed whole cells of *Lactobacillus iners* AB-1, by ProSci Incorporated (Poway, CA) following a standard two-month protocol.

### 3.2.13 Absorption of polyclonal antisera

Whole-cell lysates were prepared from overnight cultures of each of the following ten strains: *Staphylococcus aureus* subsp. Newman, *Escherichia coli* subsp. Nissle, *Atopobium vaginae*, *Gardnerella vaginalis*, *Lactobacillus rhamnosus* GR-1, *Lactobacillus reuteri* RC-14, *Lactobacillus crispatus*, *Lactobacillus gasseri* 33323, *Lactobacillus jensenii* RC-28, and *Lactobacillus johnsonii* ATCC 11506. For each culture, the cells from 1 ml of overnight culture were harvested by centrifugation, resuspended in phosphate-buffered saline (PBS), then lysed using a bead beater. Unlysed cells were subsequently removed by centrifugation and the lysates from all ten cultures were combined. Nitrocellulose membranes (5 × 5 cm squares) were incubated in the lysate for 1 h at room temperature, with the membranes being turned every 15 min. After removing the lysate and washing the membranes 3 × 10 min with tris-buffered saline (TBS), the membranes were blocked for 1 h in 10% skim milk. The polyclonal antisera was incubated with 5 blocked membranes for 60 to 90 min, with the membranes being turned over every 15 min. The membranes were replaced with 5 new membranes and incubation repeated. This process was repeated for a total of 20 membranes, with 100 μL samples being taken after each 5-membrane incubation for analysis by dot blotting.
3.2.14 Immuno-blot assay

Immuno-blot assays were performed using the Bio-Rad Goat Anti-Rabbit IgG (H+L) AP Immun-Blot® Assay Kit and the Bio-Rad Alkaline Phosphatase Conjugate Substrate Kit according to the recommendations of the manufacturer. The completeness of the absorption procedure was analyzed using the same ten strains that were used for the absorption protocol (see above) plus Lactobacillus iners AB-1.

3.2.15 Transmission electron microscopy and immunogold labeling

Overnight monocultures of L. iners grown in MRS and self-collected vaginal swabs from healthy volunteers were prepared using a standard method for electron microscopy and immunogold labeling at the Transmission Electron Microscopy Facility, Department of Microbiology and Immunology at the University of Western Ontario. After embedding samples in LR white resin and thin slicing, the sections were placed on 400 mesh nickel grids and treated with the anti-L. iners primary rabbit antisera. After washing, samples were then incubated with goat anti-rabbit IgG conjugated to 10nm gold particles (British Biocell International, EM.GAR10) for labeling. Samples were stained with 2% uranyl acetate and examined with a Philips EM 410 microscope.

3.2.16 Growth assays

A modified MRS (37) broth was made lacking carbohydrates and with the addition of bromocresol purple as a pH indicator. Carbohydrates were added at a final concentration of 20mg/ml and a final total broth volume of 10ml for each experiment. The carbohydrates tested were: glucose, lactose, mannose, fructose, maltose, and porcine stomach mucin type III (Sigma: M1778) which has a glycoprotein content similar to vaginal mucin (38, 39). Growth was compared to the standard MRS medium used to grow the strain (Difco™ Lactobacilli MRS Broth). Colony forming units (CFU) were enumerated by drop plating serial dilutions of the broth culture onto Columbia blood agar plates every 24h after inoculation.
This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession ADHG00000000. The version described in this paper is the first version, ADHG01000000.

3.3 Results and Discussion

3.3.1 Phylogeny

A phylogenetic tree was constructed from concatenated ribosomal subunit protein sequences of several Lactobacillus species available from the NCBI database with the addition of sequence for L. iners AB-1. The relationship between the species is in agreement with previous predictions (40) and places L. iners in the group known as the “acidophilus complex” (41) (Figure 3-1 and Figure S1). This clade contains two other Lactobacillus species, L. gasseri and L. crispatus, that colonize the vaginal tract, the predominantly gut microbes L. johnsonii and L. acidophilus, and one dairy culture, L. delbrueckii subsp. bulgaricus (referred to here as L. bulgaricus). L. gasseri and L. johnsonii are sister species to L. iners within this clade, although L. iners is markedly more distantly related, which indicates a divergent evolutionary step. This small clade of organisms formed the basis for genomic comparisons with L. iners due to sequence similarity and, in the case of L. crispatus and L. gasseri, the shared vaginal environment.

<table>
<thead>
<tr>
<th>Species</th>
<th>Size</th>
<th>CDS</th>
<th>HA(n)</th>
<th>HA(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. johnsonii NCC 533</td>
<td>1.99</td>
<td>1821</td>
<td>2</td>
<td>0.11</td>
</tr>
<tr>
<td>L. gasseri ATCC 33323</td>
<td>1.89</td>
<td>1755</td>
<td>2</td>
<td>0.11</td>
</tr>
<tr>
<td>L. iners AB-1</td>
<td>1.30</td>
<td>1190</td>
<td>65</td>
<td>5.46</td>
</tr>
<tr>
<td>L. crispatus JV-V01</td>
<td>2.07</td>
<td>2209</td>
<td>10</td>
<td>0.45</td>
</tr>
<tr>
<td>L. acidophilus NCFM</td>
<td>1.99</td>
<td>1862</td>
<td>4</td>
<td>0.21</td>
</tr>
<tr>
<td>L. delbrueckii subsp. bulgaricus ATCC 11842</td>
<td>1.86</td>
<td>1562</td>
<td>24</td>
<td>1.54</td>
</tr>
</tbody>
</table>

Figure 3-1. Subset of a phylogenetic tree representing L. iners as part of the acidophilus complex

Phylogeny was constructed using conserved ribosomal proteins and the default mode at phylogeny.fr (35). Key genome features are summarized for each species. The full dendrogram and detailed methodology is presented in Supporting Online Materials.
3.3.2 General genomic features

The *L. iners* AB-1 genome is the smallest *Lactobacillus* genome reported to date (11). Optical mapping demonstrated a circular chromosome of 1.304 Mbp, a size which agreed with our assembled genome scaffold of 1.301 Mbp with no plasmids (Figure 3-2 and Figure S3). Comparison of the optical map and the assembled scaffold suggests that there are 3 kb separating the two scaffolded contigs. The size of the *L. iners* genome places it within the range of several obligate symbionts and parasites (42, 43). In comparison, the smallest genome of any free-living organism is that of Pelagibacter ubique at 1,308,759 bp. The *L. iners* genome has a low GC content of 32.7% which is similar to the most closely related species *L. johnsonii* (34.0%) and *L. gasseri* (35.3%) (23). Surprisingly, the small genome still contains 6 rRNA gene operons which likely provides a competitive mechanism for responding to environmental pressures and resource availability (44). The *L. bulgaricus* genome also contains an exceptionally high number of rRNA operons relative to genome size. This was attributed to a recent phase of genome reduction (45). As *L. johnsonii* (18) and *L. gasseri* (46) each have 6 rRNA operons, the *L. iners* genome may have undergone a large genome reduction phase after splitting from the *L. johnsonii* and *L. gasseri* lineage, but the rRNA clusters were retained.
Figure 3-2. Genomic atlas of *Lactobacillus iners* AB-1

From the outer circle inward: coding regions are marked on the first two rings: outside the dividing line if encoded on the positive strand, and inside the dividing line if encoded on the negative strand. The third ring (dark green) marks ORFs predicted to be horizontally acquired. The fourth ring (orange) shows ORFs predicted to be among the top 10% most highly expressed based on CAI. The fifth ring shows local CG content measured in a sliding window as a black plot. The innermost graph shows the CG-skew with sharp changes in skew occurring at the origin and terminus of replication. Genes of interest, as described in text, are marked on the outside of the atlas. Adhesins are marked in red. The atlas was constructed using the CGView Server (47) and a large-scale version can be found in the SI Appendix (Figure S2).

The *L. iners* genome was predicted by GeneMark (20) and Glimmer (21) to encode 1190 Open Reading Frames (ORFs) greater than 100 nt with an average length of 998 nt. A large fraction of the genes (85.7%) had one or more homologs with an E-value of 1e-20 or lower in the NCBI non-redundant database. Functional classification of the 1190 predicted genes by Clusters of Orthologous Genes (COGs) (22) showed that 956 (80.3%)
were homologous to known gene families, including 201 “poorly characterized” genes. The number of pseudogenes was estimated by identifying ORFs that were truncated by at least 10% compared to the full length of the best protein match in the NCBI database (Table S1). To account for gene miscalls by GeneMark, the truncated ORFs were verified with Glimmer predictions and corrected where appropriate when Glimmer predictions produced better alignments with predicted proteins in the NCBI nrdb. The analysis resulted in 95 possible pseudogenes with approximately half predicted as hypothetical proteins or phage-related (Table S2). Notable horizontally acquired genes are described in the relevant sections following.

### 3.3.3 Comparative genomics

The InParanoid tool (33) was used to identify groups of orthologous proteins between *L. iners* and three other species of the clade (Figure 3-3). There were a predicted 766 genes shared between all four species, representing a large proportion of the *L. iners* coding sequence (64%) compared to *L. crispatus* (37%), *L. acidophilus* (44%), and *L. gasseri* (44%). This illustrates the unique nature of *L. iners* as most of its reduced genome contains core genes of the clade. Similar predictions were obtained when *L. johnsonii* was substituted for *L. gasseri* (Figure S4).
Figure 3-3. Venn diagram representing orthologous proteins between select species of the *L. iners* clade

Values are the number of orthologous genes between overlapping species as predicted by InParanoid (33). The 789 genes conserved between the four species represent a relatively large proportion of *L. iners*’ total gene content (69%).

The distribution of genes with a COG functional category was compared between *L. iners* and the four other species within the clade for which there were COG functional predictions available from the NCBI database (Figure 3-4). *L. iners* was considered to be significantly different within a category if the value was not within two standard deviations of the mean of the other four organisms. *L. iners* has significantly fewer genes in 13 of 20 COG functional categories, and it has the fewest number of genes of all five organisms in 15 categories, tying for lowest in two additional categories.

As observed previously (45), *L. bulgaricus* had approximately one half the number of carbohydrate transport and metabolism (COG category G) genes as the mean of the remaining three lactobacilli (excluding *L. iners*) creating an unusually high variance in this category. The severely reduced number of genes in *L. bulgaricus* is due to its long use as a milk starter culture, and its adaptation to this specific nutrient rich, but carbohydrate limited environment. The authors noted many instances of loss or inactivation of carbohydrate metabolizing enzymes other than lactose, in this rapidly
evolving organism. The vaginal environment is much more complex due to fluctuation of hormones affecting mucus and glycogen production, pH, and microbial species, so although L. iners also exhibited gene loss related to carbohydrate transport and metabolism, it was not to the extent of L. bulgaricus.

**Figure 3-4. Comparison of the distribution of genes by COG functional category**

The white bar represents the mean and standard deviation of the number of genes in each COG functional category for four species of the L. iners clade (L. johnsonii, L. gasseri, L. acidophilus, L. bulgaricus). The values for L. iners are plotted in the black bar. Marked with asterisks are values in L. iners that are at least two standard deviations away from the mean of the other organisms in the same COG category.

COG Functional Categories: (J) Translation, ribosomal structure and biogenesis; (K) Transcription; (L) DNA replication, recombination and repair; (D) Cell division and chromosome partitioning; (O) Posttranslational modification, protein turnover, chaperones; (M) Cell envelope biogenesis, outer membrane; (N) Cell motility and secretion; (P) Inorganic ion transport and metabolism; (T) Signal transduction mechanisms; (C) Energy production and conversion; (G) Carbohydrate transport and metabolism; (E) Amino Acid transport and metabolism; (F) Nucleotide transport and metabolism; (H) Coenzyme metabolism; (I) Lipid metabolism; (Q) Secondary metabolites biosynthesis, transport and catabolism; (R) General function prediction only; (S) Function unknown; (V) Defense mechanisms; (U) Intracellular trafficking and secretion.

### 3.3.4 Horizontal gene transfer (HGT)

Genes in L. iners and the other five lactobacilli of the same clade (Figure 3-1) were identified as foreign if the first three significant hits (E-value of 1e-20 or less) to the NCBI nr database were to a genus other than Lactobacillus. L. iners has a striking number of predicted genes likely acquired from organisms outside its genus, with 65
genes identified accounting for 5.5% of protein-coding genes (Figure 3-1 and Table S3). The next highest number was 24 foreign genes in the L. bulgaricus genome accounting for 1.5% of protein-coding genes. For the other organisms in the group, less than 1% of their protein-coding genes are of putative foreign origin. Even more striking, 26 of the foreign genes in the L. iners genome have at least 80% amino acid identity to a non-lactobacillus organism indicating a recent large-scale introduction of these genes. Many of the horizontally acquired genes are most similar to a gene found in an organism that shares the vaginal environment, possibly indicating the origin of the L. iners gene (Table S3). Interestingly, of the six members of the acidophilus group, L. iners and its sister species, L. johnsonii and L. gasseri, lack both CRISPR (clustered regularly interspaced short palindromic repeats) regions and the associated cas genes (48). Since CRISPR loci are a defense mechanism that excludes foreign DNA, this may explain, in part, why L. iners was able to acquire a large number of foreign genes. Several competence-related proteins were identified (LINAB1_1090 ComEB, LINAB1_0944 MecA, LINAB1_0943 CoiA, LINAB1_0502 ComEC, LINAB1_0501 ComEA, putative ComGC protein LINAB1_0373, LINAB1_0372 PulF, LINAB1_0371 PulE) suggesting L. iners has components of the machinery for natural competence.

It is not uncommon to find subjects colonized by one of either L. crispatus or L. iners (4, 8), suggesting there may be competition between the organisms. Pressure to compete with other organisms, including lactobacilli, may have been a reason for L. iners to acquire foreign genes, such as the adhesins and cytolysin described below. Interestingly, instillation of probiotic L. rhamnosus GR-1 with L. reuteri RC-14 or candidate probiotic L. crispatus CTV05 on its own were not inhibited by indigenous L. iners nor did they displace the L. iners from the vagina, indicating in vivo persistence of indigenous L. iners strains (1, 49).

### 3.3.5 Cholesterol-dependent cytolysin

The L. iners AB-1 genome contains a cluster of related genes making up a complete horizontally-acquired type I restriction modification (RM) system: restriction subunit (LINAB1_0222), methyltransferase subunit (LINAB1_0217), and three specificity proteins (LINAB1_0218, LINAB1_0220, LINAB1_0221). There is one additional
isolated restriction endonuclease encoded in the genome (LINAB1_1155). Neighboring the RM system is a putative integrase, and a predicted cholesterol-dependent cytolysin (LINAB1_0216) that belongs to the family of gram-positive pore-forming cholesterol-dependent cytolysins (CDC), formerly known as thiol-activated cytolysins. These proteins are typically found in pathogenic bacteria, and when expressed and secreted they use cholesterol in host cell membranes as a receptor for binding. Upon binding, a conformational change in the protein causes cell lysis by forming large pores in the membrane (50). This is the first instance of a putative cholesterol-dependent cytolysin occurring in a Lactobacillus genome.

LINAB1_0216 is 519 AA long, and has 55% amino acid identity to vaginolysin of Gardnerella vaginalis. A neighbor-joining tree derived from a multiple sequence alignment of representative CDCs shows that the putative L. iners AB-1 CDC is most closely related to the CD59-dependent intermedilysin class (Figure S5). CDC proteins have a well-conserved undecapeptide in the last domain of this four domain protein that is important for host interaction and pore formation. Hughes et al (51) recently demonstrated in Streptococcus that a short peptide sequence adjacent to the undecapeptide was responsible for specific interaction to the complement regulatory protein, CD59. Figure 3-5 shows an alignment of the undecapeptide and the CD59 interaction motif for LINAB1_0216 and its most closely related proteins. The putative cholesterol-dependent lysin was found to contain the consensus undecapeptide motif for cholesterol binding, but to lack the CD59 motif. Thus, although intermedilysin and vaginolysin use the human complement factor CD59 as a dependent co-factor with cholesterol to mediate binding (52), the sequence differences in the L. iners AB-1 lysin indicate the protein does not have the specificity to bind to CD59. We noted that domain 4 of LINAB1_0216, the membrane interaction domain, was more closely related to tetanolysins than to intermedilysins (Figures S6), supporting the conclusion that LINAB1_0216 is not dependent on CD59 for membrane binding. This putative lysin is predicted to be highly expressed by L. iners AB-1 based on a codon adaptation index (CAI) that ranks in the top 10% of this predicted measure of gene expression (Table S4).
Figure 3-5. Alignment of conserved motifs in the putative cholesterol-dependent cytolysin (LINAB1_0216) with related cytolysins

The cholesterol-binding undecapeptide extends for the first 11 residues. The residues beginning at the R in intermediysin compose the CD59 motif. The affinity of the SKN peptide for CD59 is four fold higher than is the NRT peptide. LINAB1_0216 has one or more differences from the intermediysin at each of the two positions, the most drastic being the SKN to DEQ difference that changes a polar-basic-polar trio to an acidic-acidic-polar trio.

*L. iners* is the predominant component of the vaginal microbiota in a significant proportion of healthy women (6, 53, 54). There is no clinical evidence to date that the presence of *L. iners* is associated with induction of discharge, inflammation, or discomfort. Therefore, we speculate that the cytolysin could be used to acquire nutrients from the host in a symbiotic way, or may exert antimicrobial activity against eukaryotic organisms in the environment, such as *Candida* species, or that it may contribute to attachment to host cell membranes. However, the exact function and role of LINAB1_0216 remains to be determined.

### 3.3.6 Carbohydrate transport and metabolic capabilities

Perinatally and until menopause, increased estrogen induces anaerobic metabolism of the large amounts of glycogen on the vaginal epithelium, causing an acidic environment in which the major unbound carbon sources are glucose, mannose and glucosamine (55) all of which *L. iners* AB-1 is predicted to metabolize. The vaginal mucosa consists primarily of mucin glycoproteins composed of monosaccharide chains of L-fructose, N-acetylneuraminic acid (sialic acid), galactose, N-acetyl-galactosamine, and N-acetylglucosamine (39). Growth assays (Figure S7) determined that *L. iners* AB-1 preferred carbon source is mucin with maximum growth occurring within 24-48h of
initial inoculation. Lactose, glucose, and maltose also supported growth but peaked after a longer incubation of 72-96h. We speculate that *L. iners* is adapted for an easily accessible component of the mucin polysaccharide which once depleted in culture resulted in loss of viability. In the natural vaginal environment mucin turnover (56) would provide new moieties of carbohydrates for *L. iners* to access. No growth was observed in fructose or mannose.

Complete phosphotransferase systems (PTS) were found for lactose (LINAB1_0229, LINAB1_0230 plus LINAB1_0228: LacI regulator), galactitol (LINAB1_0173, LINAB1_0174, LINAB1_0175), glucose (LINAB1_00303, LINAB1_00304, LINAB1_00306), and fructose (LINAB1_0669). The lactose-specific PTS system has greater than 75% protein identity to several *Streptococcus* strains, while being less than 50% identical to other lactobacilli. There were two complete PTS systems initially annotated as members of the mannose family of PTS transporters (LINAB1_0058 to LINAB1_0060, and LINAB1_0223 to LINAB1_0226), and one system predicted to be a mannose/fructose transporter with the most protein identity to *Streptococcus* spp. (LINAB1_0650 to LINAB1_0647). The mannose class of PTS transporters are structurally and functionally distinct from other PTS transporters, most notably for their ability to transport a wide variety of sugars including mannose, glucose, glucosamine, fructose, galactosamine and N-acetylgalactosamine (57). Two separate EIIA subunits encoded without other neighboring PTS components and have predicted specificity to mannose/fructose (LINAB1_0226) and glucitol/sorbitol (LINAB1_0935).

Clusters of ABC-type transport genes composing complete transport machinery for general sugar transport (LINAB1_0031, LINAB1_0032, LINAB1_0033, LINAB1_0035, LINAB1_0036 and LINAB1_0341, LINAB1_0339, LINAB1_0338, LINAB1_0337) and specifically for maltose (LINAB1_0077 and LINAB1_0080) were found in addition to the carbohydrate PTS systems.

Complete metabolic pathways are present to convert each transported sugar to glucose for entry into glycolysis, with the exception of galactitol. The enzyme galactitol-1-phosphate-5-dehydrogenase is missing in *L. iners* and related species *L. johnsonii* and *L. gasseri* despite predicted galactitol transport ability. Unlike *L. johnsonii* and *L.
gasseri, *L. iners* is lacking PTS systems and enzymes for sucrose and for cellobiose metabolism. A predicted exported O-sialoglycoprotein endopeptidase (LINAB1_0262) contains a conserved glycoprotease motif (Pfam: Peptidase_M22) and the ORF protein sequence has greater than 80% identity to protein sequences in other vaginally-associated lactobacilli. Four ORFs initially annotated as pullulanases (LINAB1_0206, LINAB1_0331, LINAB1_0631, LINAB1_0626) have predicted glycosidase activity due to several carbohydrate-binding and glycosidase domains. These findings of predicted carbohydrate metabolism and transport suggest that *L. iners* is adapted to hydrolyzing and extracting the carbohydrates of the vaginal mucosa.

All glycolytic enzymes are present to convert glucose to pyruvate, and from pyruvate exclusively into L-lactate, but not into other metabolites such as D-lactate, oxaloacetate, Acetyl-CoA or acetate. *L. iners* has an incomplete TCA cycle containing only fumarate reductase. Like most other lactobacilli, the organism is unable to synthesize or metabolize fatty acids.

### 3.3.7 Other transport and metabolism

Consistent with a limited metabolic capability and other members of the lactic acid bacteria, *L. iners* AB-1 dedicates a large proportion of its genome (186 (15.6%) of protein-encoding genes) to transport (Table S5). The closely related *L. johnsonii* is known for its auxotrophic nature, but in comparison only 10.7% of protein-coding genes are related to transport. Other lactobacilli in the same clade range from 4.0% for *L. bulgaricus*, and 8.0% and 9.1% for *L. gasseri* and *L. acidophilus* respectively. Unique to *L. iners* compared to *L. johnsonii* and *L. gasseri*, is a complete metal ion transport system likely transporting zinc or manganese. The system appears to be horizontally acquired and each component has from 56 to 73% amino acid identity to *Anaerococcus*, and *Finegoldia* species based on homology to proteins in the NCBI database. Manganese has long been known as an important protectant for lactobacilli against oxidative stress and has been demonstrated as a required component for growth of several *Lactobacillus* species (58, 59).
Lactobacillus iners AB-1 lacks enzymes for the biosynthesis of nearly all cofactors and vitamins. However, unlike a number of other lactobacilli including L. johnsonii and L. gasseri, L. iners AB-1 appears to have the complete biosynthesis pathway for folate with a number of the enzymes clustered together in the genome (Figure S8). It remains to be determined if folate is produced and secreted by these bacteria into the vagina.

L. iners AB-1 has the complete pathway for de-novo synthesis and salvage of pyrimidines. A partial purine metabolism pathway is present containing the enzymes required for de-novo synthesis of PRPP, but no subsequent enzymes for conversion into IMP. All genes are present for salvage of purines.

### 3.3.8 Amino acid biosynthesis

The strain is unable to synthesize any amino acids de-novo, with the possible exception of serine from pyruvate using a L-serine dehydratase (LINAB1_0514 and LINAB1_0513). Two genes allow for the interconversion of L-aspartate and L-aspargine: asparagine synthetase (asnB), and aspartate-ammonia ligase (asnA). Also present is a glutamine synthetase (LINAB1_0635) that can convert glutamate to glutamine in the presence of NH₃. Three enzymatic steps are present to catabolize L-methionine into L-homocysteine and ultimately into O-succinyl-L-homoserine, a process unique to L. iners in comparison to L. johnsonii and L. gasseri. This allows L. iners to form the methyl-donating coenzyme S-adenosyl-L-methionine (SAM) from methionine or homocysteine.

It is unknown what further role the met pathway plays in L. iners. Compensating the inability of L. iners to synthesize most amino acids, several peptidases are predicted to be highly expressed based on CAI (SI Appendix Table S4). There are also several genes involved in peptide transport, including a predicted highly expressed oligopeptide (Opp) transport system (LINAB1_0408 to LINAB1_0412).

### 3.3.9 Adherence and host interaction

The ability of lactobacilli to adhere to vaginal epithelial surfaces is believed to be important to allow colonization, host interaction, and to exclude pathogens (60). The BV-associated microbes, G. vaginalis and Atopobium vaginae, are able to form dense biofilms recalcitrant to standard metronidazole treatment (61, 62). However, L. iners AB-
1 can interfere with, and displace, *G. vaginalis* biofilms *in vitro*, suggesting a role in restoration of a healthy vagina post-BV (63).

Examination of the genome for potential adhesion factors surprisingly showed the *L. iners* genome lacks most of the known adhesion factors and conserved adhesion domains commonly found in the lactobacilli (Tables 1 and S6). Two fibronectin-binding proteins (LINAB1_0564, LINAB1_0800) and one fibrinogen-binding protein (LINAB1_0798) were found, but no mucus-binding proteins common to lactobacilli that colonize the gastrointestinal tract were identified (64, 65), thus suggesting that *L. iners* uses different adherence mechanisms in the vagina.

The predicted ORFs were evaluated for cell-wall anchor motifs that indicate the protein is exposed on the surface of the cell. Sortase-dependent proteins (SDPs) often have a role in bacterial-host interactions and adhesion (66), and many gram-positive cell-wall anchored proteins contain a conserved C-terminal LPXTG motif that is cleaved by the sortase A enzyme between the threonine and glycine residues. The cleaved protein is then covalently linked to the peptidoglycan of the cell wall (67). Recently, the probiotic *L. rhamnosus* GG was shown to have genes for three secreted LPXTG-like pilins (spaCBA) and a pilin-dedicated sortase that mediates binding to mucus (28).

Bereft of pili (Figure 3-6), one ORF of *L. iners* was identified as sortase A (LINAB1_0482), and Table 1 shows that 6 ORFs were identified as possible SDPs containing the sortase cleavage motif LPXTG (Pfam ID PF00746). Two of the proteins (LINAB1_0273, LINAB1_0795) have several tandemly repeated Rib domains (PF08428), which may function as adhesion and biofilm factors to human epithelial cells (68, 69). All the SDPs have multiply repeated domains of unknown function within the protein which we speculate are involved in adhesion.
Figure 3-6. Transmission electron micrograph of immunogold labeled *L. iners* AB-1 (black arrows) in association with human vaginal epithelial cells

Vaginal swab samples collected from healthy women were prepared for EM by thin-sectioning. The *L. iners* cells were labeled with a polyclonal antiserum raised in rabbits against formalin-fixed whole cells of *L. iners* followed by a secondary goat anti-rabbit gold-conjugated antibody (10 nm gold particles). An even surface distribution of gold particles is expected for *L. iners* but not other bacteria because antibodies not specific to *L. iners* AB-1 were absorbed out of the serum as described in the Supplemental Text. The scale bar represents 500 nm.

A large fraction of *Streptococcus* and *Staphylococcus* SDP surface proteins with the LPXTG anchor motif also contain the YSIRK signal motif (70, 71). Deletion and substitution mutations of the YSIRK motif in cell-wall-anchored protein A of *Staphylococcus* reduced secretion and sorting of the protein. Except for LINAB1_0273, all of the LPXTG proteins identified (Table 1) in *L. iners* also contained the YSIRK motif (PF04650) and another two proteins had the N-terminal YSIRK-type signal motif without an associated LPXTG motif. The combinational signal and anchor motifs in these proteins suggest they are under tight regulation to be properly secreted and anchored to the cell surface.
In addition, several of the putative adhesion factors show weak similarity to adhesion proteins. Two of the five SDPs have a significant hit (E-value of 1e-20 or less) to the NCBI nr database: LINAB1_0273 and LINAB1_0701 are similar (26 and 47% respectively) to a possible biofilm-associated protein of BV pathogen *Atopobium vaginae*. Six putative binding proteins had significant hits to the nr database, however, only two of these had greater than 50% amino acid identity. LINAB1_0796 had 65% identity to a possible mucus binding protein precursor of *L. jensenii*. However, LINAB1_0796 itself does not contain a mucus-binding domain, and there is no significant similarity to any other mucus-binding protein in the database suggesting that the *L. jensenii* protein is mis-annotated.

To determine if any putative cell-surface proteins were expressed, cell-wall protein extract from *L. iners* was separated by SDS-PAGE electrophoresis and analyzed by Western Blot with antisera raised against *L. iners* AB-1. Bands were isolated for mass spectrometry peptide mass fingerprinting. Isolated proteins were putatively identified based on predicted size in kDa (Figure S10) and the peptide mass coverage. One band (marked with a black arrowhead in Figure S10) was predicted to be LINAB1_0950 based on 6 matched peptides, and 15% peptide mass coverage. When search options included variable lysine acetylation, 16 peptides covered 25% of the protein. The 48294 Da predicted mass of this protein is consistent with the size of the band. The protein sequence contains a YSIRK signal sequence supporting its predicted location outside the cell, but has no significant similarity to proteins in the NCBI non-redundant database.

In order to further confirm the localization of the isolated proteins to the cell surface, transmission electron microscopy (TEM) of *L. iners* monoculture was prepared with a rabbit polyclonal antisera was raised against formalin-fixed whole cells of *L. iners* and labeled with IgG conjugated to 10 nm gold particles. This showed specific labeling distributed around the cell membrane (Figure 3-6 and Figure S9). Vaginal swab samples collected from healthy women were confirmed by PCR to contain *L. iners* (72) and TEM and immunogold-labeling using the *L. iners* antisera, showed *L. iners* to be consistently closely associated with vaginal epithelial cell surfaces (Figure 3-6).
The lack of proteins containing adhesion domains and several cell-anchored proteins lacking sequence conservation to known proteins suggest that *L. iners* has a unique and unknown method of adherence. In support of this assertion, 7 of the 14 putative adhesion proteins have a CAI greater than the mean CAI, and two are predicted to have gene expression levels among the highest 10% of all genes. Taken together, these observations suggest that these genes may provide a competitive advantage in its specialized environment, but further characterization is required.

### 3.3.10 Stress tolerance and environmental response

*L. iners* AB-1 has the ability to survive both the low pH (3.6 to 4.5) conditions of a healthy vagina and the higher pH (greater than 4.5) conditions associated with BV (73, 74). The genome contains two alkaline-shock proteins (LINAB1_0403 and LINAB1_0434), also present in *L. crispatus*, which may aid in pH tolerance of the vaginal environment. The genome contains a heat-shock operon with a heat-inducible transcriptional repressor, HrcA, and molecular chaperones GrpE, DnaK, and DnaJ, plus a second stress-induced operon controlled by a redox-sensing transcriptional repressor and the chaperonins GroES and GroEL. Two universal stress proteins (LINAB1_0760 and LINAB1_0227), and a cold-shock protein (LINAB1_0419) were identified. Collectively, these may aid in allowing the organism to tolerate different pHs, mucus changes, and stress associated with infection.

RNA polymerase contains four standard subunits and an additional sigma factor that directs the enzyme to a specific promoter. Alternative sigma factors are active under different stress conditions to regulate the transcription of various stress response genes. In addition to the typical RpoD (LINAB1_0602) sigma factor, *L. iners* has three alternative sigma factors identified as RpoE (LINAB1_0946), RpoH (LINAB1_1028), and RpoN (LINAB1_0645). Interestingly, RpoE is used by the opportunistic intracellular pathogen *Burkholderia cenocepacia* for growth under stress (75). Another intracellular Gram negative pathogen *Brucella melitensis* uses RpoH and RpoN to cope with stress from elevated temperature and exposure to H$_2$O$_2$ (76). These factors may contribute to the ability of *L. iners* to survive in the presence of H$_2$O$_2$-producing strains in the vagina, and oxidative stresses associated with infection. LINAB1_0645 (RpoN) is adjacent to a
complete mannose/fructose PTS system with high amino acid identity to strains of *S. pyogenes*. Upstream is an unknown conserved protein (LINAB1_0651) with an RpoN interaction domain. Recent work by Stevens et al. (77) characterized an RpoN-dependent mannose PTS system of *Lactobacillus plantarum* with a similar operon structure and RpoN acted is a major regulator of carbohydrate uptake.

There were four complete two-component systems identified, each consisting of a histidine kinase and a response regulator. Three of the paired systems have high similarity to proteins from other lactobacilli, while one system (LINAB1_0792, LINAB1_0791) appears to be horizontally acquired with 50-60% amino acid identity to proteins from either *Anaerococcus prevotii* or *Oribacterium sinus*. In addition to the complete coupled systems, three more response regulators without a corresponding histidine kinase were found (LINAB1_0366, LINAB1_0143, LINAB1_0785).

We noted that a cluster of genes shared between *L. iners*, *L. crispatus*, and *L. johnsonii* encode an iron-sulfur (Fe-S) cluster assembly system (LINAB1_0712 to LINAB1_0717). We examined homologs in the NCBI database and found these proteins to be most similar to these lactobacilli known to colonize the vaginal tract, but the complete system was rare in lactobacilli inhabiting other niches. Three of the *L. iners’* proteins in the cluster (LINAB1_0712, LINAB1_0714, LINAB1_0716) did not match any of the proteins in the NCBI non-redundant database at greater than 60% protein identity, suggesting they have a specified or unique function in *L. iners*.

The Fe-S clusters were first characterized as proteins needed to carry iron and sulfur to appropriate areas of the cell and as electron transfer molecules (78). The so called Suf system in *E. coli* is activated under conditions of limiting iron or oxidative stress, such as the presence of H$_2$O$_2$ (79). Because of the sensitivity of a Fe-S cluster to oxygen, they can act as environmental sensors and regulate a response by the organism (80). This could be an important mechanism for resistance to oxidative stress in *L. iners* where an abundance of H$_2$O$_2$ is produced by other lactobacilli in the vagina (81). Although iron is a required element for nearly all living organisms, it has commonly been accepted that lactobacilli do not require iron for growth and instead use manganese and cobalt as cofactors for
biological processes (58, 82). However, Elli et al. (83) found that depleting iron from a nucleotide limited growth media resulted in no growth of certain *Lactobacillus* species, while Duhutrel et al. (84) reported the acquisition and use of iron in *Lactobacillus sakei*, an organism especially well adapted to meat. There were no apparent iron uptake systems in *L. iners*, but a ferrochelatase (LINAB1_0037) was detected and is not present in the other species of the clade. This enzyme plus an oxygen-independent coproporphyrinogen III oxidase (hemN) could break down heme for transport (85), although no transport system has yet been characterized and the hemN protein is lacking in *L. iners*. The potential for *L. iners* to sequester iron in the Fe-S complex could limit the availability of iron to vaginal pathogens, and provide a competitive advantage against organisms unable to use iron in this iron-rich environment.

Finally, we noted that all steps in the oxidative branch of the pentose phosphate pathway are present for the production of NADPH, an important molecule for the prevention of oxidative stress.

### 3.3.11 Defense mechanisms

The vagina harbors a highly diverse community of competing microbes. Bacteriocins are produced by vaginal lactobacilli and enterococci (86, 87) and the ability to resist their antimicrobial effects would be advantageous for *L. iners* survival. No complete bacteriocin synthesis genes were found in the *L. iners* AB-1 genome, but three genes were found to encode putative bacteriocin immunity proteins. LINAB1_0852, identified by RAST as a microcin self-immunity protein, has 91.5% amino acid identity to a putative bacteriocin immunity protein of *Finegoldia magna* and greater than 60% identity to MccC microcin immunity family protein from various strains of *S. pneumoniae*. Resistance to microcin may be significant as it is an antimicrobial peptide produced by uropathogenic *E. coli* associated with aerobic vaginitis and urinary tract infection (88). This along with a bacitracin resistance protein (LINAB1_0937), and putative enteroicin A immunity protein supports the assertion that *L. iners* incorporates foreign genes that may contribute to its ability to resist vaginal eradication by these pathogens. Of note, *L. iners* AB-1 has an additional putative pore-forming hemolysin gene (LINAB1_0545) of *Lactobacillus* origin in addition to the horizontally-acquired cholesterol-dependent
cytolysin described previously. Three other proteins contain domains of the hemolytic-related Tyl family (89, 90) (LINAB1_0440, LINAB1_0684, LINAB1_0698). The hemolytic potential of L. iners speculatively may be used for defense, or may play a role in providing nutrients and surviving through menstruation. Interestingly, L. iners AB-1 exhibits clearing of sheep’s red blood cells in culture (Figures S11), but it is unclear whether the putative hemolysins or cholesterol-dependent cytolysin are responsible.

3.4 Concluding remarks

This study describes the smallest Lactobacillus genome reported to date, with a genome that appears highly adapted for a specialized niche environment. The genome contains the largest fraction of genes composed of conserved proteins within the clade, and it also has the largest fraction, and absolute number, of genes predicted to be acquired from foreign sources. The conserved genes are largely core metabolic proteins shared between lactobacilli. L. iners is present in the vagina that is deemed healthy, infected with BV, or has just been subjected to antimicrobial therapy (5, 7). Recent quantification of bacterial numbers in these different stages of vaginal health has demonstrated that the abundance of L. iners remains relatively constant despite fluctuating environmental conditions (9). This remarkable ability to survive under a range of conditions suggests that rather than L. iners being somehow associated with an aberrant microbiota, it may be an important member of the host’s defenses by being a persistent commensal lactobacilli involved in restoration and maintenance of the normal microbiota. Further analysis of the functional characteristics associated with the L. iners’ genome may uncover desirable characteristics of a commensal microbe that contributes to the maintenance, and potentially restoration, of a healthy and stable vaginal microbiota.

3.5 References


19. R Development Core Team (2013) R: A Language and Environment for Statistical Computing


86. Dezwaan DC, et al (2007) Purification and characterization of enterocin 62-6, a two-
peptide bacteriocin produced by a vaginal strain of *Enterococcus faecium*: Potential

antimicrobial peptide lactocin 160, produced by the vaginal *Lactobacillus rhamnosus* .

88. Azpiroz MF, Poey ME & Lavina M (2009) Microcins and urovirulence in

beta-hemolysin gene of *Brachyspira* (Serpulina) hyodysenteriae and its expression in

Chapter 4

4 Comparative meta-RNA-seq of the vaginal microbiota and differential expression by *Lactobacillus iners* in health and dysbiosis

This chapter is reproduced with permission (Appendix C) from:


Supporting material for this chapter including Tables and Figures (labeled as S#) and supporting Material and Methods are available as a PDF online:

http://www.microbiomejournal.com/content/supplementary/2049-2618-1-12-s1.pdf

4.1 Introduction

Studies using high-throughput metagenomic and 16S rRNA sequencing have identified over 250 bacterial types in the human vagina. Microbial profiles in women who are clinically healthy most often have a low diversity microbiota dominated by lactobacilli with the most common species being *Lactobacillus crispatus, Lactobacillus iners, Lactobacillus jensenii,* and *Lactobacillus gasseri* (1-4). Conversely, for bacterial vaginosis (BV), an aberrant condition associated with increased risk of sexually transmitted infections and preterm labor (5, 6), several high-diversity, multi-species profiles have been reported (3, 7). This makes the human vaginal microbiota different than other human microbial ecosystems where a high species diversity correlates with healthy conditions. For example and in contrast, inflammatory bowel disease is characterized by a loss of diversity (8). The vaginal microbiota is highly dynamic and bacterial populations can change rapidly between the healthy and BV states (1, 4, 9), but the cause for these transitions is unknown. The most frequently detected organism,
*Lactobacillus iners*, appears to have a streamlined genome adapted for persistence in the vagina (10). This organism is detected in women regardless of BV status (7, 11), but not much is known about how *L. iners* can adapt to these differing environments or its role in the etiology and pathogenesis of BV.

Although there are numerous 16S rRNA studies of the vaginal microbiome that analyze the relative microbial composition (e.g. (1, 3, 12)), none have yet attempted to characterize the function of the microbiota using culture-independent methods. We therefore chose a meta-transcriptomic approach using RNA-seq to address the functional contribution of the bacterial community. We further sought to understand the function of *Lactobacillus iners*, an organism found nearly ubiquitously in the vagina.

### 4.2 Materials and Methods

#### 4.2.1 Clinical samples

Premenopausal women between the ages of 18-40 years were recruited at the Victoria Family Medical Center in London, Canada. The Health Sciences Research Ethics Board at the University of Western Ontario granted ethical approval for the study under HSREB #16183E. Participants were excluded from the study if they reached menopause, had a urogenital infection other than BV in the past 6 months, were pregnant, had a history of gonorrhoaeae, chlamydia, estrogen-dependent neoplasia, abnormal renal function or pyelonephritis, were taking prednisone, immunosuppresives or antimicrobial medication, or had undiagnosed abnormal vaginal bleeding. Participants were asked to refrain from oral or vaginal intercourse and consuming probiotic supplements or foods for 48 hours prior to the clinical visit. No participants were menstruating at time of the clinical visit. After reviewing details of the study, participants gave their signed informed consent before the start of the study. Vaginal swabs were collected from four women: two with BV and two considered to have a non-BV vaginal microbiota (N) as diagnosed by Nugent scoring (13), and vaginal pH (Table S1). A nurse obtained vaginal samples for RNA-seq using a Dacron polyester-tipped swab rolled against the mid-vaginal wall and immediately suspended in RNAprotect (Qiagen) containing 100 ug/ml rifampicin. A second swab collected in the same way was smeared onto a slide and air dried for Nugent
scoring (13) of bacterial vaginosis (Nugent scoring is presented in Table S1). Vaginal pH was measured using the pHem-alert applicator (Gynex). Samples for RNA extraction were incubated at room temperature for at least 10 minutes (to a maximum of 3 hours), and then centrifuged before discarding the supernatant and freezing the remaining pellet at 80°C. Lysis and RNA extraction were performed within 3 weeks of storage.

4.2.2 RNA isolation, mRNA enrichment, and sequencing

Cell pellets were lysed in a 700 ul solution of 20 mg/mL lysozyme and 50 U/mL mutanolysin for 20 minutes at 37°C with periodic vortexing. After lysis, the samples were centrifuged (5500 x g for 15 minutes) and the supernatant was discarded. The remaining pellet was used for RNA extraction by TRIzol (Invitrogen) according to the manufacturer’s protocol. After RNA isolation by TRIzol, 8-9 ug of total RNA was used for rRNA depletion using a single round of MICROBExpress (Ambion) according the manufacturer’s protocol. RNA quality and subsequent rRNA depletion was verified before and after the MICROBExpress treatment by Agilent 2100 Bioanalyzer (Bioanalyzer results before and after MICROBExpress treatment shown in Figure S1). Samples were DNase treated with the TURBO DNA-free kit (Ambion). Samples were sent to the Toronto Center for Applied Genomics (TCAG, Toronto, Ontario, Canada) for library preparation and sequencing by ABI SOLiD 4.

4.2.3 Reference sequence library and mapping

A total of 110 accessions representing 103 organisms (of 31 genera, and 63 species) isolated from or detected in the vagina were included in a reference sequence set for mapping (Table S4). The dataset represents the partial or complete genomes available from the NCBI database as of March 2011. These 234,991 sequences (including 230,031 coding sequences) were clustered by sequence identity (95% nucleotide identity over 90% sequence length) using CD-HIT-EST (14) to remove redundancy in the reference mapping set. A representative sequence (“refseq”) from each of the resulting 163,014 clusters was used to build a Bowtie (15) colorspace reference library for mapping the RNA-seq reads. Reads mapped uniquely by Bowtie to a coding refseq were included in the differential expression analysis (all other unmapped reads were discarded). Reads
were trimmed from the 3’ end to 40 nt, and up to 3 mismatches were allowed. Reads with equal best hits were mapped at random to one of the locations.

### 4.2.4 Functional assignment of refseqs

SOLiD colorspace format precludes direct functional assignment of sequenced reads, and therefore annotations were assigned to the refseqs following mapping. Amino acid translations of predicted coding sequences were compared to the Clusters of Orthologous Groups of proteins (COGs) database (16) by rps-blast using an e-value cutoff of 1e-3 to assign a COG function. SEED Subsystems (17) were similarly assigned by blastp with an e-value cutoff of 1e-3. The KEGG Automatic Annotation Server (KAAS) (18) was used for annotation of enzyme functions (KOs) and mapping of metabolic pathways. All database comparisons and annotations were performed on the data available as of April 2011.

### 4.2.5 Statistical analyses for differential expression

We used the ANOVA-Like Differential Expression (ALDEx) R package version 1.3.0 (19) for differential expression analysis. Refseqs belonging to L. iners were used to evaluate differential gene expression at the single organism level. For differential expression of the meta bacterial community we modified the ALDEx package to group refseqs into higher functional levels (SEED subsystem4, and KEGG KO assignments). For genes or functional groups to be considered differentially expressed, we required a log₂ relative difference (ΔR) of at least 2 for function-level analysis and at least 1.5 for L. iners gene-level analysis. To be considered differential we also required that ΔA_q ≤ 0.01, implying that less than 1% of the values in distributions between conditions overlap. A brief description of the statistical framework is in the supplemental materials and presented in Figure S9.

### 4.2.6 CRISPR spacer analysis

CRISPR spacer sequences were downloaded from the CRISPRs database (http://crispr.u-psud.fr/crispr/) and RNA-seq reads were mapped to these sequences as described for refseq mapping.
4.2.7 cpn60 reference mapping

Sequences belonging to the cpn60 gene were downloaded from cpnDB (20) and were manually curated to remove redundant sequences (sequences with 100% identity to another sequence in the database) so that one representative sequence remained. RNA-seq reads were mapped to this reference set as described for refseq mapping.

4.3 Results and Discussion

Vaginal swabs from four women were collected for RNA-seq of the bacterial transcriptome. Two women were diagnosed with BV and two with a non-BV (or “Normal”) vaginal profile according to Nugent score, vaginal pH, and signs and symptoms of the condition as noted by the examining clinician (see Table S1 for Nugent scoring). Sequenced reads were mapped against a reference coding sequence library (“refseqs” - See Methods) resulting in 5,487,128 to 10,635,713 uniquely mapped reads per sample; far exceeding the sequencing depth of recent RNA-seq studies (21-23). The number of unique refseqs mapped ranged from 10,770 to 22,860 per sample; with more refseqs expectantly identified in BV samples due to the higher microbial diversity (mapping is summarized in Table S2). Among all four samples, 33,412 unique refseqs were identified in total. To verify the completeness of our reference library and ensure no major known taxon was missing from analysis, we performed an independent mapping against a reference cpn60 gene database. The highly expressed cpn60 (or groEL) gene is nearly universally conserved in bacteria and contains variable sequence which allows for taxonomic discrimination (20), and unlike the rRNA molecules, is not depleted by the mRNA enrichment process. Mapped cpn60 reads representing at least 0.1% of the total cpn60 mapped reads are presented in Table S3. The data suggested that our reference library contained representative genomes for the most abundant organisms detected in our samples, and that any unavailable genomes would not make up an appreciable fraction of the bacterial or mRNA population.

The fraction of mapped reads per taxon for each sample is represented in Figure 4-1A. The microbiota associated with health, samples N4 and N30, were dominated by lactobacilli (L. iners and L. crispatus) while those associated with BV, B27 and B31,
were a mixed population composed of lactobacilli and anaerobic organisms that included *Prevotella, Gardnerella,* and *Megasphaera.* The profiles are consistent with those from healthy and BV subjects studied by others (1-4), and corresponded to the Nugent profile for each sample (Table S1).

---

**Figure 4-1. Bacterial community structure and function**

**(A)** The distribution of all mapped mRNA reads for each taxon per sample. The two samples that were considered healthy (N4 and N30) are dominated by *Lactobacillus crispatus* with lesser amounts of *L. iners* and *L. jensenii* detected. The two samples having bacterial vaginosis (BV) have a lower relative proportion of lactobacilli and large proportions of *Prevotella amnii, Gardnerella vaginalis, Megasphaera,* and other anaerobic organisms. **(B)** Selected subsystem 2 and subsystem 3 categories showing enriched subsystem 4 functions under N or BV conditions. Each point represents a specific subysys4 function, and the magnitude of change in expression in plotted on the x-axis relative to the BV condition. Points are colored blue (non-BV) or red (BV) if significantly differentially expressed between conditions whereas grey points are not differential between the conditions. **(C)** Representation of the fractional taxon
contribution to functions significantly upregulated in BV (red points in B). Pairs of bars represent the two samples (B27 and B31) and each color is the fraction of reads for that function contributed by a particular taxon.

Butyrate kinase, as part of the “Acetyl-CoA fermentation to Butyrate” subsystem, is highlighted to show the relationship between the differential functions in 1B and the taxonomic composition of the function in 1C.


The multi-organism composition of our samples, and the resulting gene diversity, violates assumptions of current differential RNA-seq analysis methods. These methods generally perform well when used to study the transcriptome of single organisms, however they implicitly assume that the expected between-condition difference is essentially a fixed effect (24, 25). Since each sample in this study was obtained from a different subject, a random-effect-type model is more appropriate. However, it can be problematic to use this model when there are not enough samples to accurately estimate the additional parameter(s) required by random-effect models. We therefore developed a specialized statistical framework termed “ANOVA-Like Differential Expression” or “ALDEx” (19) that infers differential expression by estimating the magnitude of between-condition expression difference with respect to within-condition expression difference. Thus genes (or functional groups) identified as differentially expressed between the N and BV conditions have a between-condition (N vs. BV) expression difference that is large in comparison to the within-condition (B27+B31 and N4+N30) variability. Furthermore, since we posit that biochemical function, not organism is the major unit of analysis for the vaginal microbiome the data were grouped via refseqs into functional groups (KO or SEED subsys4) unless otherwise noted. Functional-level analyses of metagenomic data, as opposed to gene-level, has been used previously (26). Unsupervised hierarchical clustering showed that samples were more similar when refseqs were grouped by function (Figure S2B) than clustering by refseq expression (Figure S2A). We also noted, regardless of refseq grouping, the two BV samples were most similar to each other based on gene expression and were separable from the two N samples that also clustered together. Estimated between vs. within expression ratios along with read counts and annotations are presented in Table S8 (inners refseqs), Table S9 (subsys4), and Table S10 (KEGG KOs).
4.3.1  *Lactobacillus iners'* response to BV

*Lactobacillus iners* is the most frequently detected vaginal organism, and is known for its consistent presence in N and BV states (1, 3, 27). We detected *L. iners* in our samples with an average coding sequence (CDS) coverage per sample ranging from approximately 5-fold to 117-fold (Table S5). This allowed us to uncover gene expression differences in this organism between the two vaginal conditions.

We mapped a total of 1653 unique *L. iners*’ CDS with at least one read and refer to this set as the *L. iners* “pan-transcriptome” (Figure 4-2). There were 954 of the 1653 CDS present in all 12 *L. iners* genomes that represent the “core” expressed CDS set. The core functions, in general, had a high relative expression and low fold-change between conditions (Figure 4-2). A total of 207 CDS were inferred to be differentially regulated due to their large estimated between vs. within expression ratios. These ratios, along with read counts and annotations, are presented in Table S8.
RNA-seq reads were mapped to 12 available *L. iners*’ genomes (listed in Table S3) after clustering redundant coding sequences by nucleotide identity (see Materials and Methods). Breaks in the circle separate blocks of contiguous coding sequences (CDS) ordered by scaffolding on the genome assemblies (Ring ①). The height of the plot represents how many of the 12 genomes the CDS occurs. The first contiguous block (dark blue) are the 954 CDS present in all genomes and are considered the “core” *L. iners* gene set. Ring ② shows the COG color-coded function of each CDS. Ring ③ The differential expression of each CDS between Non-BV and BV samples. The absolute height of the bar shows the fold-change (log$_2$) in expression (positive for up BV, and negative for up non-BV). Bars are colored if significantly differential (red for
BV and blue for non-BV). Regions of interest are labeled in the centre of the circle. Rings 4 and 5 show a heatmap representation of the median CDS expression in BV and Non-BV respectively. Darker blue represents higher expression. Regions of interest noted in the results are marked by lettered triangles.

Amongst the most highly differentiated CDS were eight contiguous genes encoding the proteins for the CRISPR anti-bacteriophage defense system and include the 5-member (CasA-CasE) protein complex, Cascade (CRISPR-associated complex for antiviral defense) (28). These CRISPR-related genes were highly expressed only under BV conditions (Figure 4-1BC and Figure 4-2) suggesting that L. iners is responding to an altered environmental phage load. Interesting, the cas proteins were only present in 4/12 L. iners genomes (Figure 4-2) which raises the question of strain-specific adaptations to the vaginal environment that could not be identified by 16S rRNA profiling. It is also possible that these differences reflect different genome assembly qualities. We noted that restriction-modification systems were also highly upregulated by L. iners in BV (Figure S3), perhaps as further defense against bacteriophage infection.

The cas proteins, as part of the CRISPR anti-bacteriophage system, regulate the insertion and presentation of small pieces of DNA spacer sequences in the bacterial chromosome used to inactivate attacking phage DNA (29). We therefore probed our RNA-seq data for evidence of CRISPR spacer sequences and detected between 21,477 and 225,242 potential CRISPR spacers per sample of which the vast majority corresponded to known sequences of Lactobacillus origin (Table S6). Examining the meta-RNAseq data also revealed 6 highly upregulated CRISPR-associated proteins (Figure 4-1B). Most of the reads were contributed by L. iners (as described above), but there was shared expression of these functions by Gardnerella vaginalis and Megasphaera, and one of the cas proteins was expressed only by Prevotella species (Figure 4-1C). Our observations are supported by recent evidence that several Lactobacillus-specific bacteriophage have been isolated from women in South Africa with and without BV (30), and that metagenomic data from the Human Microbiome Project has also revealed a diversity of CRISPRs in every body site with an abundance of Lactobacillus-targeting phage sequence in the vagina (31).
A notable core CDS, a cholesterol-dependent cytolysin (CDC) encoded by *L. iners* (10), was upregulated at least 6-fold in BV (Figure 4-2B). This CDC is active between pH 4.5 and 6 (32) which corresponds to the elevated pH range of BV (33). The predicted protein is similar to vaginolysin, another CDC present in *Gardnerella vaginalis* that is upregulated at least 256-fold in the meta-RNA-seq analysis (Table S9). Vaginolysin has been shown to have cytotoxic activity towards human erythrocytes, vaginal epithelial, and cervical cells (34). This raises the possibility that the *L. iners*-encoded CDC may play an unappreciated role in BV and might contribute to the pathogenesis of the condition.

Examination of the distribution of expressed reads assigned to COG functions showed an increase in transcriptional effort by *L. iners* for Carbohydrate Transport and Metabolism and a number of individual CDS under this category were upregulated in BV (Figure S4). In the genomic context, three separate loci involved in carbohydrate uptake were upregulated 3- to 9-fold in BV (Figure 4-2: two loci are marked in the region labeled A and the third by C). The first is a broad-specificity phosphoenolpyruvate-dependent transport system (PTS) of the mannose-family along with a putative regulator of the mannose operon, ManO, possibly targeting carbohydrate moieties of the vaginal mucosa (10). This PTS is conserved in all 12 genomes of *L. iners* suggesting its functional importance to the organism in the vagina. A second region, also conserved across all sequenced *L. iners* genomes, contains an ABC-type maltose transport system next to a maltose phosphorylase suggesting a preference for maltose uptake under BV. The source of maltose could be from breakdown of glycogen by *L. iners*, and consistent with this, there are four upregulated glycosylases predicted to target α-1,6-glucosidic linkages that bridge the branching points in glycogen. The third locus related to carbohydrate utilization contains a second mannose-family PTS flanked by an oligo-1,6-glucosidase de-branching enzyme and a LacI family transcriptional regulator. This region is present in only 6 of the 12 available *L. iners* genomes and suggests that only a subset of strains are able to use these genes for adaptation during BV.

Though glycogen is thought to be the major carbon source for vaginal bacteria (35), we surprisingly noted under BV conditions the up-regulation of a tandem set of three genes related to glycerol metabolism including a glycerol kinase, a glycerol-3-phosphate
dehydrogenase, and a glycerol facilitator (glpF) (region marked “D” in Figure 4-2). The genes together suggest that *L. iners* is able to uptake glycerol for conversion to glycerol-3-phosphate and then glycerone phosphate for entry into glycolysis or glycerophospholipid metabolism.

We examined the glycerol utilization genes in the context of the meta-transcriptome (Figure 4-1B and C). The glpF glycerol transporter and downstream glycerol kinase are highly expressed by both *L. iners* and *L. crispatus* and therefore results in no differential expression of these genes at the community level. In addition to the glpF transporter, *L. iners* and *L. crispatus* express a UGP-type glycerol-3P transporter more strongly under N conditions. *Prevotella amnii*, a BV-associated organism, also expressed an independent glycerol-3P uptake system (glpT) (Table S9). Bypassing the ATP-using phosphorylation of glycerol by glycerol kinase, glpT and ugpB/E allows direct uptake of glycerol-3-phosphate which the organism can use for phospholipid biosynthesis or conversion to glycerone-P using glycerol-3-phosphate dehydrogenase (expressed highly in BV) for entry into glycolysis. Two enzymes (Glycerol-3-phosphate cytidylyltransferase and Glycerol-3-phosphate O-acyltransferase) are expressed highly under N conditions by *L. iners* and *L. crispatus* and indicate glycerol-3P is also converted into products for lipoteichoic acid and glycerophospholipid synthesis. The overall high expression and the differential transcription of these glycerol-related genes suggests glycerol is an underappreciated molecule in the vaginal bacterial ecosystem.

As an additional evaluation of gene expression by *L. iners*, we collected and sequenced mRNA from a single *L. iners* strain, AB-1, grown to mid-log phase in MRS broth (Difco, BD). Figure S10 and Table S8 summarize the mapping and gene expression for this sample. Similar to the *in vitro* expression data, core genes were highly expressed *in vitro* and many potentially strain-specific genes were not detected. Regions for carbohydrate uptake and metabolism (marked A, C, and D in Figure 4-1 and Figure S10) were amongst the most highly expressed. One notable difference between the *in vitro* expression compared to the clinical samples was the lack of CRISPR expression (region E in Figure 4-1A and Figure S10). Also notable, the cytolysin is highly expressed under culture conditions. These data show that genes that may be key for *L. iners’* survival in the
vagina are induced under *in vitro* conditions opening possibilities for future experiments to test gene expression and the responding phenotype.

### 4.3.2 The vaginal community in BV

As described in Materials and Methods, we grouped refseqs (representing genes from one or multiple organisms) by their annotated function in order to evaluate functional differences between N and BV that may be contributed by multiple organisms. Using SEED Subsystem annotations, we estimated expression differences for subsys level 4 functions using ALDEx (output is presented in Table S9). The expression of subsys4 level functions were visually represented in two ways: 1) As strip plots where the subsys4 functions were plotted by absolute fold-change between conditions (Figure 4-1B, and Figures S5ABC) colored by significance and 2) as heatmaps where the summed expression value for each subsys4 was averaged over the number genes per function (Figures S6ABC). The strip plot allows us to visualize under which functional category the subsys4 functions are binned, while the heatmaps demonstrate the transcriptional effort by the community for each function based on the subsys4 expression relative to the average expression of all subsys4. In addition to identifying differential functions, we can show the taxonomic composition of the function (Figure 4-1C and Figures S7 and S8) allowing us to discern which organisms are producing transcripts for each function.

At the broadest functional level (SEED subsystem 1), the most highly expressed subsys4 functions in both conditions belong to general cell maintenance and information processing functions such as Cell Division and Cell Cycle, Nucleosides and Nucleotides, and RNA and DNA metabolism (Figure S6A). However, many of the differentially expressed subsys4 functions for these categories were more highly expressed under the N condition (Figure S5A). The trend was conserved at the more specific SEED subsystems level 2 and level 3 (Figures S5BC and S6BC). This management of cell growth and proliferation is consistent with known constraints on apportioning cellular resources (36). We suggest that some of the transcriptional resources are necessarily diverted from housekeeping tasks to other functions in BV because of the greater diversity of organisms with a wider array of metabolic capabilities.
Similar to the findings for *L. iners*, several functions related to carbohydrate metabolism were found to differ between N and BV at the SEED subsystem 2 and 3 levels (Summarized in Figure 4-1B and Figure S5BC). The BV samples were enriched at subsystem 2 functional categories of Glycoside hydrolases, and Polysaccharides, and also at subsystem 3 in Cellulosome, and Glycogen metabolism. The functions belonging to these categories are enzymes involved in uptake, degradation, and metabolism of glycans. Considering the context of the vagina, these glycan targeting genes are likely used for the metabolism of glycogen which accumulates in the vaginal epithelium (37). The genes involved in these functions largely belong to *Prevotella amnii* and some to *Gardnerella vaginalis* (Table S7). In addition to several glycosidases and glycosyltransferases, there are components of an outer membrane-bound “starch utilization system” (Sus), belonging to *Prevotella amnii*, which are expressed in BV and have been implicated in binding and transporting of starch by several intestinal bacteria (38). Interestingly, the upregulation of a “starch phosphorylase” gene by *Prevotella amnii* indicates it may be able to synthesize glycogen via UDP-glucose; a phenomenon observed in other *Prevotella* isolated from the ruminal gut (39). This could allow *Prevotella* to control carbohydrate resources by storing excess for a time when it might be needed due to growth needs, or limits in environmental carbon sources. Supporting glycogen as an important carbon source for the BV organisms is the up-regulation of oligo-1,6-glucosidase and glucokinase: the enzymes responsible for directing glycogen into glycolysis via alpha-D-glucose.

Similar to the SEED subsystem analysis, we next grouped the refseqs by KEGG KO numbers for differential expression analysis. We used the KO and EC assignments to place differential functions in corresponding KEGG pathways to examine the community gene functions in the context of linked metabolic pathways. Examined this way, the two conditions differed primarily in energy metabolism. The major steps of glycolysis and the conversion of pyruvate to lactic acid was upregulated under N indicating that lactic acid was the main metabolic end-product; consistent with the predominance of lactobacilli in these samples. In contrast, for the BV state SEED subsystem 2 (Figure 4-1B and Figure S5B) showed a number of upregulated functions in electron donating reactions belonging to steps of the TCA cycle and oxidative phosphorylation leading to the production of succinate. A high ratio of succinate to lactic acid was the original marker of BV (40, 41),
and high succinate concentrations distinguish BV from the similar but distinct condition of aerobic vaginitis (AV) (42). Increasing succinate at the expense of lactate results in a self-perpetuating condition whereby an elevated pH and a more reduced environment (43) supports the growth of the anaerobic bacteria noted in BV. In addition to succinate, we observed an up-regulation of butyryl-CoA dehydrogenase and butyrate kinase leading to butyrate production from acetyl-CoA under BV conditions. Examination of these functions linked to butyrate production in SEED subsystems showed these were expressed largely by *Prevotella amnii* and by *Megasphaera*. Butyryl-CoA dehydrogenase and butyrate kinase are shown as first and second enzymes in the “Acetyl-CoA to Butyrate” function in Figure 4-1C. These short-chain fatty acids have been shown to modulate immune function in the vagina and these anti-inflammatory properties may contribute to the non-inflammatory nature of BV in contrast to the inflammatory aerobic vaginosis (AV) condition (42, 44, 45).

One of the distinctive characteristics of symptomatic BV is the “fishy” odor upon application of 10% KOH attributed to increased polyamines in the vaginal fluid produced by amino acid breakdown (46). *Prevotella amnii* expressed an arginine decarboxylase under BV conditions which could facilitate the synthesis of putrescine from arginine. Interestingly, we also noted a spermidine synthase gene expressed by *Dialister* and *Megasphaera* in BV which can convert putrescine to another polyamine, spermidine. Although *Dialister* is a minor constituent of our samples, it has been detected in vaginal studies and associated with diverse microbiota profiles indicative of BV (1, 7). This could indicate that production of odor is a cooperative process depending on the particular microbiota composition. Indeed, odor is not universally detected in women who are BV-positive by Nugent (47), and previous studies suggest that particular taxa are more associated with odor (3, 48). Symbiotic relationships may be common in the vaginal environment as another has been described where *Gardnerella vaginalis* produces amino acids that promote the growth of *Prevotella bivia* (49).

### 4.4 Concluding Remarks

We used a novel method to characterize the conserved differences between the metatranscriptome of two healthy and two BV microbial communities. Despite having four
distinct microbial populations, our results show that each condition displays a set of conserved metabolic capabilities that, while influenced by the organisms present, is largely conserved between samples of differing structure. A summary of major findings is presented in Figure 4-3. The predicted metabolic end-products are known to produce the acidic or strongly reduced environments associated with health or BV respectively. Surprisingly, glycerol may have an underappreciated role in this environment with expression profiles suggesting upregulation of pathways to use glycerol as a carbon source under BV conditions. The source of glycerol is unknown, but it is an abundant component of mammalian and bacterial cell membranes, and extracellular glycerol levels are used in clinical practice to measure loss of cellular integrity. Thus, the rupture of cells by cytolysins expressed by *L. iners* and *G. vaginalis*, as well as lipase activity could release glycerol and glycerol components into the environment.
The predominant lactobacilli present in the healthy condition results in a relatively simple environment where carbohydrates are converted to lactic acid. This results in a low pH and inhibition of the growth of other anaerobic organisms. The more complex microbial composition in BV results in an increase in cell lysis, and an increase in carbohydrate availability. In this case, the carbohydrates are converted to succinate and other short chain fatty acids that can modulate host immunity, increase the pH and increase the reducing environment. Concomitant with this is an increase in bacteriophage load, and the potential for several organisms to co-operate to produce the malodorous compounds found in some cases of BV.

The BV-associated organism *Prevotella amnii* showed enhanced glycogen extraction during BV, as well as the ability to synthesize its own glycogen. We posit a possible symbiotic relationship between the BV-associated organisms in our samples for the production of polyamines responsible for vaginal odor during BV.

Figure 4-3. Overview of predicted differential functions of the vaginal microbiota based on RNA-seq analysis
We additionally found that the vaginal microbiota is responding to bacteriophage populations differently in healthy and BV conditions. Several longitudinal studies show rapid unexplained shifts in the vaginal microbiota (4, 9), and one study suggested that phage have a role in the depletion of lactobacilli during episodes of BV (50). This idea is supported by a recent study of viruses in the human fecal microbiota which proposed that phage populations may be markers of change in the bacterial community (51).

We describe a different expression profile of Lactobacillus iners in health and BV. The persistence of L. iners under these drastically differing environments (different pH, different bacterial population composition, and different bacterial load) could be due to its ability to respond and regulate its genomic functions. We suggest some of these regulated functions, such as specific carbohydrate uptake, bacteriophage defense, and a cytolysin, could be adapted for survival during episodes of BV. Some expressed functions were attributed to a subset of the L. iners genomes. Strain-specific functions of the microbiota, often overlooked by current 16S rRNA sequencing studies, can be extracted using function-based sequencing. Further investigation into the role of L. iners as a passive or active participant in the etiology and pathology of BV could help us understand this highly prevalent condition.

Overall, examining the meta-vaginal transcriptome has highlighted the complexity of this environment. The large difference in bacterial function between the two conditions emphasizes our need to better understand the interactions between species in the context of the rest of the community and of the host. It is likely that examination of additional microbiota types associated with health and BV will identify similar broad patterns, but will provide additional insights into the association between host symptoms and the action and products of the microbiome.

4.5 References


Chapter 5

5  Effect of antimicrobial and probiotic therapy on the vaginal microbiota

This chapter is based on a manuscript under editorial consideration for publication at Science Translational Medicine.

Supporting Table S5-1 is available for download as an additional file (.xlsx)

5.1  Introduction

The vaginal microbiome plays an important role in health and disease. Two of the most common medical problems for women are bacterial vaginosis (BV) and vulvovaginal candidiasis (VVC) (1, 2). Therapy for these conditions has changed little in over 40 years, and recurrences of both conditions are extremely common, reducing quality of life for many women (3-5). The antimicrobial failures are likely due to drug resistance, failure of the agents to penetrate and eradicate biofilms, and inability to restore bacterial homeostasis. Our recent clinical studies reported that adjunctive oral use probiotic *Lactobacillus rhamnosus* GR-1 and *Lactobacillus reuteri* RC-14 for 28 days with 2g single dose of tinidazole resulted in a higher cure rate (87.5% by Amsel criteria and 75.0% by Nugent criteria) versus 50.0% (Amsel) and 34.4% (Nugent) with tinidazole and placebo (6). The same dosage of probiotics and single dose of fluconazole (150 mg) improved the cure rate of VVC (defined as having no vaginal discharge, itching and/or burning vaginal feeling, dyspareunia and/or dysuria, and Candida recovered by culturing) from 10% to 35% compared to fluconazole treatment alone (7).

We hypothesized that the probiotics therapy helped restore a *Lactobacillus*-dominated microbiota, based upon previous clinical and *in vitro* studies (8, 9). We investigated this by performing targeted sequencing of the V6 region of the 16S rRNA gene from the vaginal samples we had collected before and after the tinidazole therapy in 62 premenopausal women with BV. In the case of VVC, the condition is not associated with depletion of lactobacilli or an abnormal microbiota profile per se (10), so the aim of the
sequencing of samples from 55 women was to confirm these previous findings and generate a testable hypothesis that would explain why probiotics can increase the cure rate for this condition.

5.2 Materials and Methods

5.2.1 Clinical samples and study design

Vaginal swabs were collected from Brazilian women as part of previous clinical studies (6, 11, 12), and were stored at -80°C until DNA extraction for amplification and sequencing. The samples were obtained from subjects diagnosed with bacterial vaginosis (BV) (N=62) by Amsel (13) and Nugent (14) criteria before treatment, and a second sample was collected four weeks following randomized treatment. Patients were randomized into one of two treatment groups, either probiotic or placebo. For both groups, treatment consisted of a single 2g dose of tinidazole plus either two oral capsules of L. rhamnosus GR-1 and L. reuteri RC-14 or two placebo capsules taken daily in the morning, for twenty-eight days starting on the first day of tinidazole use as per the published study (6). The placebo group received two gelatin capsules containing cellulose and magnesium stearate that were also taken daily in the morning for twenty-eight days starting on the first day of tinidazole treatment. Due to failure of PCR amplification, we dropped 2 patients (one from each treatment group) from the dataset compared to the original study by Martinez et al., 2009 (6) resulting in 62 patients in this study compared to 64 originally.

In a separate but similar study, vaginal swabs were obtained from women diagnosed with vulvovaginal candidiasis (VVC) (N=55) by positive Candida and signs of vaginal irritation and discharge and excluding women having BV as assessed by Nugent scoring. Women were randomized into two groups where both groups received a single 150 mg dose of fluconazole (an anti-fungal). Women in the probiotic group were also administered two oral capsules of L. rhamnosus GR-1 and L. reuteri RC-14 taken daily in the morning for twenty-eight days starting on the first day of fluconazole use (7). The placebo group received two gelatin capsules containing cellulose and magnesium stearate that were also taken daily in the morning for twenty-eight days starting on the first day of
fluenceoazole treatment. Swabs for biota profiling were collected at the end of the 28-day period. Single time point vaginal swabs were also collected from a population of healthy, age-matched Brazilian women (N=63) with no signs or symptoms of vaginal irritations or infections, and with a low, or “normal”, Nugent score of 0-3.

Each subject signed an informed consent under a protocol (Protocol No. 0146) approved by the Ethics Review Board at the CSE-FMRP-USP. This study was registered online at Comissão Nacional de Ética em Pesquisa (CONEP; Document No. 070202), Brazil. The details of the clinical study design are described in the original study paper by Martinez et al. (6) as follows:

“Identical vials containing probiotics and placebo were received from the manufacturer in separated boxes with 2 different colors and they were random numbered, by staff not participating in the study, in the laboratory at Universidade de São Paulo, Brazil. Investigators were blinded to the randomization code until all data were analyzed. Equal parts of probiotics and placebo vials were placed in plastic bags and given to investigators to take to their health centers and randomly distribute to patients enrolled in the study. The subjects were treated on a firstcome basis with a single dose of tinidazole (2 g) plus either 2 oral capsules of L. rhamnosus GR-1 and L. reuteri RC-14 or 2 placebo capsules taken daily in the morning for 28 days, starting on the first day of tinidazole use. Assuming an anticipated cure rate of 60% in the control group arm, a minimum of 25 subjects per group was required to detect an increase in the cure rate as compared with 90% in the group supplemented with the probiotic microorganisms, assessed at the 2-sided 5% level of significance with 80% power.

After 4 weeks of treatment, the same procedure adopted during the first appointment with the gynecologist was performed. The presence of any symptoms and signs (vaginal discharge, fishy odor, or burning and itching vaginal feeling) or possible side effect related to the drug and (or) probiotic were recorded. The investigators remained blinded to the study codes until all analyses had been completed.”
5.2.2 V6-targeted 16S rRNA gene sequencing

Swabs were vigorously shaken in 1mL PBS (pH 7.5) to dislodge cell material. DNA extraction was performed using the InstaGene Matrix (Bio-rad) according to manufacturer instructions. Amplification of the V6 region of the bacterial 16S rRNA gene was performed as described previously (15), using unique barcode sequences embedded in the V6 primers for each sample. The V6 region was chosen specifically due to length restrictions on the Illumina GAIIx, which was the platform available to us at the time of sequencing, and for the ability to easily distinguish *Lactobacillus* and *Gardnerella* species (15, 16). The amplified products were pooled for Illumina GAIIx 100bp paired-end sequencing at the Next Generation Sequencing Facility at The Centre for Applied Genomics (TCAG) at the Hospital for Sick Children in Toronto, Canada.

5.2.3 Sequence processing

Paired reads were overlapped using Xorro (https://sourceforge.net/projects/xorro-overlap/) (17) to give full-length V6 sequence. Sequences with at least 97% identity were clustered into operational taxonomic units (OTUs) using UClust v. 3.0.617 (18). The most abundant sequence in each OTU cluster was used for taxonomic assignment by RDP SeqMatch (19). From the top 20 matches to the RDP named isolates database, the full taxonomy was retained for matches with the highest S_ab score. For multiple top matches with equal scores, the lowest common ancestor in the taxonomy was retained (e.g. genus level if multiple species matched equally well). Taxonomy assignment was verified by BLAST (20) to a custom V6 vaginal organism database. Up to 1 mismatch was accepted as a hit. The read counts table and OTU taxonomy are presented in a Supplementary Data File: Table S5-1.

5.2.4 Data analysis

Weighted UniFrac distances (21) between samples were calculated in QIIME (22) by using a phylogenetic tree of OTU sequences built with FastTree (23). The phylogeny was based on an OTU sequence alignment with MUSCLE (24) since alignment to a consensus model was not possible with the short hypervariable V6 sequences. Shannon’s diversity was calculated for each sample in QIIME. Figures were generated and statistics
were performed in R (25). Welch’s two-sample t-test was used to compare UniFrac distances before and after treatment between groups.

5.3 Results

5.3.1 The normal vaginal microbiota in pre-menopausal Brazilian women

The healthy, asymptomatic vaginal samples set the baseline for the expected bacterial profiles in the population (Figure 5-1). Most of 63 samples showed a dominance of *Lactobacillus* spp. with 44/63 samples (69.9%) having at least 60% of the sequences belonging to the *Lactobacillus* genus. *Lactobacillus iners* and *L. crispatus* were the predominant species. As expected, this population had a low Shannon diversity (Figure 5-2) because the microbiota composition is dominated by a few abundant organisms. After *Lactobacillus*, *Gardnerella* spp. was the next most abundant genus, and along with *Prevotella* and *Megasphaera* sequences, were abundant in samples with < 60% *Lactobacillus*. 
**Figure 5-1. Genus-level bar plot of vaginal bacterial abundance measured by V6 sequencing in 63 healthy, asymptomatic Brazilian women**

*Lactobacillus iners* and *L. crispatus* are plotted at the species level. Since sequences belonging to *Lactobacillus* are plotted first (from the bottom). The average-linkage dendrogram at the top represents the weighted UniFrac distance between samples. The major clusters defined by the major branch points are samples dominated by *L. crispatus*, *L. iners*, or a mixture of organisms. Most samples (44/63 = 69.9%) are dominated by at least 60% *Lactobacillus* sequences. The color legend shows the top 14 most abundant taxa and applies to Figure 5-1, Figure 5-3, and Figure 5-6.
5.3.2 Vulvovaginal candidiasis (VVC) study

As shown in Figure 5-3 (top row) women with VVC generally had *Lactobacillus*-dominated profiles (40/55 = 72.7% with >= 60% *Lactobacillus* spp.) with *L. iners* and *L. crispatus* being the most predominant species. The distribution of profiles was similar to those in the healthy, asymptomatic group. The bacterial diversity in the VVC group before treatment, as measured by Shannon’s diversity index in Figure 5-2, was similar to that of healthy women (median value of 1.76 in VVC and 1.88 in healthy).

![Figure 5-2. Distribution of Shannon diversity](image)

The Shannon diversity index was calculated for all biota samples at time 0 (before treatment) plotted by group. The bacterial diversity of women with BV was much higher than the healthy women or women with VVC.

The UniFrac phylogenetic distance metric was used to measure the phylogenetic distance between each before/after treatment pair in the placebo and control groups. A lower UniFrac score indicates more similarity between the compared biota with a score of 0 representing biota profiles exactly the same in bacterial presence and abundance, while a
score of 1 represents profiles with no overlapping bacterial sequences (21). Figure 5-4 shows box plots of the paired UniFrac distance measures for the placebo and probiotic treatment groups.

![Box plots of paired UniFrac distance measures](image)

**Figure 5-3.** Genus-level bar plot of vaginal bacterial abundance measured by V6 sequencing in women diagnosed with vulvovaginal candidiasis (VVC)

*Lactobacillus iners* and *L. crispatus* are plotted at the species level. Since sequences belonging to *Lactobacillus* are plotted first (from the bottom) the dashed line at the 0.6 fraction represents 60% *Lactobacillus* abundance in the samples. Like the healthy asymptomatic women (Figure 5-1) most profiles show a dominance of lactobacilli before treatment (top row). Women were randomized into two groups to receive fluconazole plus either a placebo (“placebo” group) or oral probiotic *L. reuteri* RC-14 + *L. rhamnosus* GR-1 (“probiotic” group) with the biota after treatment plotted on the bottom row. Patient biota before and after treatment are vertically aligned and patients are plotted in the same order. The taxonomy figure legend is presented as part of Figure 5-1.

After treatment with fluconazole and probiotic or placebo, there was little change in the bacterial microbiota as measured by the UniFrac distance. The low UniFrac distance between paired samples before and after treatment (probiotic group median = 0.08, placebo group median = 0.17) indicates there is high similarity in the bacterial profiles before and after treatment. Additionally, there was no significant difference between the
probiotic and placebo groups (t = -1.5189, df = 49.13, p-value = 0.1351 and 95% confidence interval (CI) = -0.22 to 0.03). Plotting the change in relative abundance for the top most abundant organisms also showed there was little shift in either the placebo or probiotic group (Figure 5-5). We also examined the UniFrac distances between all pairs of samples before treatment vs. between all other samples and noted that the median distance between samples of different individuals was greater (probiotic = 0.26, placebo = 0.31) than the distance between the same individual before and after treatment (probiotic = 0.08, placebo = 0.17). This indicates there was less intrapersonal variation in the bacterial profiles before and after treatment than interpersonal variation in the population.

Figure 5-4. Paired UniFrac distance between individual patients before and after treatment for BV or VVC

The weighted UniFrac distance was calculated for each sample pair before and after treatment as described in the Materials and Methods. There is no significant difference between treatment groups in the VVC study. A significant difference in UniFrac distance was found between the placebo and probiotic treated BV groups (p
= 0.0009187 by Welch’s two sample t-test) suggesting a larger shift in the microbiota occurred in the probiotic group compared to the placebo group.

Figure 5-5. Violin plots showing the change in bacterial abundance within patients before and after treatment

The top 6 most abundant genera are plotted for BV (top row) and VVC (bottom row). Organism colors match the bar plots (Figure 5-1, Figure 5-3, Figure 5-6) and legend (Figure 5-1). The white dot represents the median, and the thicker black line is the range from the lower to upper quartile. The thickness of the plot represents the density of points at that value. The plots show there is little change in abundance following treatment for both VVC groups and BV treated with tinidazole + placebo, but an increase in *Lactobacillus* abundance in the probiotic-treated BV group at the expense of *Gardnerella vaginalis*, *Prevotella*, *Megasphaera*, *Atopobium*, and *Sneathia* spp.

5.3.3 Bacterial vaginosis (BV) study

As shown in Figure 5-6 top row, the initial BV microbiota profiles were similar to those of other women with BV as previously reported (26, 27). In general, *Lactobacillus*-dominated profiles were rare compared to the healthy and VVC groups with only 6/62 = 9.7% women having >= 60% *Lactobacillus* sequences. The BV group had far more
bacterial diversity (as measured by Shannon diversity) than the VVC or healthy populations (Figure 5-2).

![Image](image_url)

**Figure 5-6. Genus-level bar plot of vaginal bacterial abundance measured by V6 sequencing in women diagnosed with bacterial vaginosis (BV)**

*Lactobacillus iners* and *L. crispatus* are plotted at the species level. Since sequences belonging to *Lactobacillus* are plotted first (from the bottom) the dashed line at the 0.6 fraction represents 60% *Lactobacillus* abundance in the samples. Profiles before treatment show a lack of lactobacilli and an abundance of *Gardnerella vaginalis*, *Prevotella*, *Megasphaera*, and *Atopobium*. Women had been randomized into two groups to receive tinidazole plus either a placebo (“placebo” group) or oral probiotic *L. reuteri* RC-14 + *L. rhamnosus* GR-1 (“probiotic” group) with the biota after treatment plotted on the bottom row. Patient biota before and after treatment are vertically aligned and patients are plotted in the same order. The symptomology as measured by Amsel and Nugent are indicated in the heat map below the bar plots. The top row is presence (dark grey)/absence (white) of pH > 4.5, 2nd row is presence/absence of fishy odor from application of KOH, 3rd row is presence/absence of milky white discharge. The final row represents the Nugent status (BV = dark grey, intermediate = light grey, Normal = white). The taxonomy figure legend is presented as part of Figure 5-1.

As with the VVC group, we analyzed the change in microbiota by measuring the UniFrac distances between paired samples before and after treatment (Figure 5-4). We found that
the variation in the initial time point for the treatment groups was similar with a median UniFrac distance of 0.35 for both the probiotic and placebo group and an interquartile range (IQR) of 0.29 and 0.31 respectively. We observed that the tinidazole with placebo treatment had a very small effect on the microbiota composition relative to the initial composition, with a median UniFrac distance of 0.22 with an IQR of 0.18. In contrast, the tinidazole with probiotic treatment resulted in a large shift in UniFrac distance with a mean distance of 0.57 with an IQR of 0.52. The range of values in the two treatment groups were distinct with a mean difference of 0.22, 95% CI of 0.09 to 0.36, indicating a large effect of the probiotic treatment. Because each sample contributed to only one UniFrac distance (before/after treatment for the same subject), and the placebo and control groups were non-overlapping, the usual concerns about lack of statistical independence in sets of distances isolated from a distance matrix do not apply here. We therefore used a Welch’s two sample t-test to test the hypothesis that the placebo and probiotic treatments resulted in equivalent shifts in the microbiota, and observed that they were statistically different, p-value = 0.0009187 (t = 3.4216, df = 56.934). We conclude that the probiotic plus tinidazole treatment had a significantly larger effect in altering the microbiota composition than tinidazole treatment alone, and led to the final communities closely resembling the healthy communities (Figure 5-1).

Examination of the taxonomic bar plots (Figure 5-6) suggests that the major reason for the large shift in the microbiota was the acquisition of lactobacilli in many members of the probiotic cohort, and specifically L. iners. The change in abundance can be seen more easily in Figure 5-5 top row, where the relative change of the most abundant organisms is plotted. Here it is obvious that the relative abundance of L. iners, and L. crispatus, increased in the probiotic cohort after treatment at the expense of G. vaginalis, Prevotella spp., Megasphaera spp., Atopobium spp., and Sneathia spp. In contrast, there is very little change in these organisms in the placebo group. Correspondingly, after treatment the median Shannon’s diversity was lower in the probiotic group (median = 2.15) compared to the placebo group (median = 2.78).

Clinical signs of BV were measured before and after the intervention by Nugent (14) and Amsel (13) criteria. The Nugent scoring system involved a Gram-stained vaginal smear
where bacterial morphotypes presumable *Lactobacillus, Gardnerella*, and *Mobiluncus* are enumerated. The scores are weighted heavily on the number of lactobacilli and not surprisingly a low or “normal” Nugent score corresponded with recovery of *Lactobacillus* abundance after treatment (Figure 5-6, bottom). Amsel criteria indicate BV if three of four criteria are present: pH greater than 4.5, presence of thin, white, homogenous discharge, a positive “whiff test” - the release of fishy odor after addition of potassium hydroxide, and the presence of “clue cells” (epithelial cells covered in bacteria) by microscopy. Like the Nugent score, presence of the Amsel criteria generally corresponded with *Lactobacillus*-depleted biota (Figure 5-6, below colored taxonomic bar plots).

Of note, the administered probiotics *L. rhamnosus* and *L. reuteri* were not detected at the relative abundance cutoff of 1% after treatment indicating that in these 61 women the probiotic strains did not become dominant colonizers of the vagina following one month's treatment.

### 5.4 Discussion

The relative V6 profiles for the VVC study emphasizes that this condition is not associated with a disruption of the bacterial microbiota and a depletion of lactobacilli. These organisms were highly abundant both during VVC and after anti-fungal therapy (Figure 5-3). Although we did not target sequencing for the yeast, the previous report of the clinical study showed a significant reduction in yeast following therapy (7), and the bacterial profiles show a lower bacterial diversity compared to the BV groups which is more representative of the healthy biotas (Figure 5-2). It has been suggested that amines produced by BV organisms can inhibit *Candida* and prevent concomitant yeast infection (28), but the reverse has not been investigated where *Candida* might inhibit BV organisms. *Candida* are known to produce a number of metabolites that might be inhibitory to bacteria (29), but none would explain why the Gram-negative bacteria appeared inhibited not the Gram-positive lactobacilli. In this clinical cohort, BV was an entry exclusion, so the outcome of VVC therapy on concomitant BV could not be evaluated.
If the bacterial profiles *per se* did not influence the *Candida* and yet the cure rate was improved by probiotic use compared to antifungal alone (7), what mechanism of action might be involved? A recent study showed that *Candida albicans* lost metabolic activity and eventually were killed in the presence of *L. rhamnosus* GR-1 (30). Furthermore, a genome-wide transcriptional analyses showed significantly lower expression of the *CDR1* gene encoding an important drug efflux pump involved in fluconazole resistance, suggesting the *Candida* may not regard the lactobacilli or lactic acid as a threat to viability, and this might explain why conjoint probiotic therapy increased eradication of the fungi. The lactobacilli repressed the *PHR1* and *ALS12* genes involved in biofilm formation, and increased expression of stress-related genes. In addition, oral probiotic therapy with the two strains used here has been shown to decrease transfer of yeast from the rectum, suggesting another means of stopping the reseeding process (31).

It is known that oral administration of *L. rhamnosus* GR-1 and *L. reuteri* RC-14 can restore health to patients with BV (31-33), but the present study demonstrated that this treatment also restores an indigenous lactobacilli-dominated biota. Our results support a recent study (34) showing restoration of *Lactobacillus*-dominated profiles after treatment for BV with intravaginal metronidazole and probiotic *L. delbrueckii* subsp. *lactis* DM8909. In our study, the significant tenfold increase in subjects with dominant lactobacilli compared with only twofold change with antibiotic and placebo reiterates that the current gold standard of antibiotics to treat BV is not sufficient to restore bacterial homeostasis in the vagina.

The sequence data showed some discrepancy with the clinical outcome reported previously for BV (6). Examination of the Nugent scores after treatment showed that 11/31 patients (35.5%) of the placebo group and 23/31 patients (74.2%) of the probiotic group had a normal *Lactobacillus*-dominated Gram stain after treatment. Examination of these *Lactobacillus*-dominated samples by sequencing showed that *Lactobacillus* was not the predominant organism in every case. Only 7/11 = 63.6% (placebo) and 21/23 = 91.3% (probiotic) of those previously reported as cured by Nugent had at least 50% relative abundance of lactobacilli sequences (Figure 5-6). This discrepancy may be caused by the different protocols: Nugent scoring counts absolute numbers of cell
morphotypes visible on select microscopic fields, while amplicon sequencing determines the relative abundance of lactobacilli compared to the rest of the sampled sequences and not absolute numbers.

These findings also show that even without colonizing the vagina in high numbers, orally administered lactobacilli can still influence the vaginal ecosystem. This may be through lowering pathogen ascension from the vagina, disrupting the BV biofilms (9) or promoting recovery of the indigenous lactobacilli via immune modulating effects (35) or by affecting the mRNA expression of the community (36). The cause of BV has not been established, but as with other microbial ecosystems an exogenous disturbance could lead to ecosystem degradation allowing a number of species to grow and outcompete the typically healthy lactobacilli. By administering probiotics the system is again disturbed but in a way that displaces the BV-associated organisms and allow the indigenous species to re-establish, ideally creating a stable microbiota (37).

5.5 Concluding Remarks

In this study we have shown that oral tinidazole and probiotic *Lactobacillus reuteri* RC-14 + *L. rhamnosus* GR-1 restores the indigenous vaginal lactobacilli in women with BV resulting in a better cure rate than tinidazole treatment alone. Additionally, this is the first high-throughput 16S rRNA gene sequence study of the vaginal microbiota in relation to active *vulvovaginal candidiasis* (38). Despite an apparent lack of change in the bacterial microbiota structure and abundance, oral probiotics with fluconazole resulted in better cure of VVC compared to fluconazole alone. The results underscore our lack of fully understanding the interactions between the vaginal microbiota, other pathogens, and the host immunity and environment.

5.6 References


25. R Development Core Team (2013) R: A Language and Environment for Statistical Computing


Chapter 6

6 General discussion

Extending back to 1683 when Antonie van Leeuwenhoek observed bacteria for the first time from the scrapings of his teeth (1) there has been a fascination with the microbial world associated with the human body. Though important for understanding function, culturing approaches don’t allow for the entirety of the microbial ecology to be examined at once. A number of species are inaccessible because we haven’t yet been able to adequately create a favourable environment in vitro to culture them. High-throughput sequencing has changed the course of microbiology and our understanding of the influence the microbes have on our health. The human vaginal microbiome has long been of interest for its impact on reproductive and sexual health.

6.1 Defining normal

It’s apparent from studies presented in this thesis and by others in the field that in the majority of reproductive-aged women the healthy vaginal microbiota is dominated by Lactobacillus (2). We have found that for Brazilian (Chapter 5) and African women with HIV (Section 1.4.2) the healthy biota had a predominance of L. iners that was detected in every sample. In other cases, healthy biota had a dominance of L. crispatus and rarely did another Lactobacillus species dominate. These findings are common in other studies we’ve examined of asymptomatic, healthy Canadian women (data not published). However, as exemplified by the 63 Brazilian women without signs or symptoms of vaginal abnormalities, not every individual is dominated by Lactobacillus, and some have a biota with increased bacterial diversity. Using a 60% Lactobacillus abundance cutoff, we reported that 30% of these healthy women had a mixed-species vaginal profile. The taxa associated with non-Lactobacillus-dominant healthy profiles included organisms traditionally associated with BV like G. vaginalis, Prevotella, Megasphaera, Atopobium, Leptotrichia. This represents a large inter-personal variation in the apparently healthy vaginal microbiota. Without longitudinal sampling it is unclear whether these diversified profiles are stable or possibly represent a risk for these women to develop BV. Ravel et
al. (3) reported a cohort of asymptomatic women who had a multi-species microbiological profile, and these were more likely to be ethnically Hispanic and black. Although the aim of the work presented in this thesis was not to compare ethnic differences, subjects from three continents were examined (unlike Ravel's study of only North Americans), and the findings help us better address the clinical dilemma of how to define an abnormal vaginal microbiota. The 'gold standard' Nugent score for evaluating BV is entirely based on the premise of a normal vaginal microbiota being dominated by *Lactobacillus*. If Brazilian and African women represent cohorts where mixed-species biota are more prevalent, then Nugent scoring could over-diagnose BV. Indeed, in women of African descent there is a higher report of so-called 'asymptomatic' BV (4). Possibly there are differences in genetics, changes in the local immunity or environment, or cultural practices and socio-economic factors, or the biota is less stable leading to higher risk of BV, AV, or infectious conditions.

For post-menopausal women, “normal” is less clear. Previous studies report that post-menopausal women have a lower prevalence of *Lactobacillus* due to lacking glycogen produced by the vaginal epithelium (5). However, this appears an over-simplification, and indeed lactobacilli are clearly present in such women (Heinemann et al; Devillard et al.). In Chapter 1, partial 16S rRNA sequencing revealed an association between the microbiota composition and the severity of vaginal dryness and atrophy in menopausal women. In fact, the biota were relatively stable over 10 weeks and the symptoms did not vary considerably (Figure 2-3). Dryness was rated on a four-point scale as none, mild, moderate, and severe. Six of the seven women who experienced only mild or no vaginal dryness had profiles dominated by *Lactobacillus iners* or *L. crispatus*. Of the nine women who experienced moderate or severe dryness, only one maintained a *Lactobacillus*-dominated profile of >60% relative abundance for the 10 weeks. This supports the view that lactobacilli abundance still plays a role in health at menopause, despite the physiological changes that have occurred. Although we didn’t measure glycogen levels, a microarray analysis of the vaginal epithelial transcriptome indicated genes involved in maintenance of the epithelial cell barrier were dysregulated in the moderate/severe dryness group. Although cause and effect are not determinable, the depleted glycogen and lower abundance of lactobacilli are associated with aberrant vaginal health.
Another cause of illness and discomfort in post-menopausal women is UTIs, and this can be associated with and increase the risk of urinary incontinence (6). It’s been proposed that the persistent colonization of *E. coli* in post-menopausal UTI is due to the vaginal absence of protective *Lactobacillus* species (7). This is supported by the finding that *Lactobacillus* were absent by culture in 93 post-menopausal women with a history of recurrent UTI, but were recoverable in 61% of the 36 women who received intravaginal estriol and none of the 24 receiving placebo (8). The estriol group also had a significant drop in pH and *Enterobacteriaceae* colonization while the placebo group was unchanged for both. This indicated the potential to restore the lactobacilli population as a means of reducing recurrence of UTI. Not only did estrogen therapy achieve this, but also the administration of probiotic *L. rhamnosus* GR-1 and *L. reuteri* RC-14 (9, 10).

There is a clear indication that estrogen has a modulating effect on the vaginal microbiota of post-menopausal women. A study of the temporal variability of in reproductive-aged women over a 16-week period suggested there were large changes in the community associated with points in the menstrual cycle (11), confirming earlier work by Keane *et al.* (12). Clearly, more studies are needed to understand the role of host factors in temporal dynamics of the vaginal microbiota clearly need further study, particularly if we are to differentiate a 'normal' or 'healthy' status to one that requires intervention.

It is now evident that 'normal' can have many contexts and it represents a network of interactions between the host and the resident microbes. Once inter-personal variations and the factors associated are better understood, as well as the variations within an individual over time, it will be easier to differentiate the normal fluctuations from high-risk profiles.

### 6.2 Bacterial vaginosis

The data presented in Chapter 5 and Figure 1-1 show a diverse array of vaginal profiles during clinically diagnosed (Amsel and Nugent) BV. We previously reported (13) the profiles of 272 vaginal samples clustered into distinct groups representing the dominating organisms. For BV, there were four groups represented by *Prevotella bivia*, BVAB1 (previously reported *Lachnospiraceae*), *Gardnerella vaginalis*, and one group of mixed
species. The profiles from the Brazilian BV cohort in Chapter 5 had a similar array of dominating BV organisms. The distribution of the measured diversity as calculated by Shannon’s diversity index for both studies had a median value of ~2 for non-BV/healthy, and ~3 for BV (Figure 5-2) indicating the increased complexity of the microbiota during BV. These findings agree with those reported by others, namely that health and BV are not simple conditions with single end-points of bacterial species.

Being able to survey the vaginal microbiota at sequencing depth allows one to investigate the question of a so-called “rare microbiome”, and how many species are present in small numbers in an environment. In a large dataset of women with HIV and BV we determined there were approximately 60 OTUs, and the 272 samples we studied had a median sequence depth of >40,000 reads per sample (13). For the 323 samples composing the data in Chapter 5, the average sequencing depth was >38,000 reads per sample and we reported 77 OTUs. These studies far exceed the sequencing depth of other vaginal microbiota studies. For example, the 396 samples described by Ravel et al. (3) were sequenced at an average depth of 2,000 reads per sample and reported 282 taxa and Srinivasan et al. (14) reported a median 1,779 reads for 220 samples identifying 796 species. The discrepancy in the reported number of OTUs could be explained by the difference in sequencing platform and the accumulation of errors. The pyrosequencing method used by Ravel et al. and others has a notoriously high error rate resulting in spurious reads divergent from the original sequence population (15). In addition, amplifying multiple variable regions across a constant region (e.g. V1-V2, or V3-V4 used in the studies mentioned above) creates more opportunity for chimeric reads comprising sequence fragments from two different genomes (16). In our Illumina datasets, the dominant error originated from PCR amplification and chimeric sequences, and other sequence errors were exceedingly low and did not contribute to the segregation of OTU clusters (17). Measures of species richness with ACE and Chao1 methods and rarefaction curves supported sufficient depth of sampling to reach near saturation. Regardless of the number of rare organisms in the vagina, the top abundant organisms we report in health and BV have a high congruence with those reported by other groups (3) using different methodologies and sequencing platforms. This suggests we have a relatively
comprehensive picture of the “who’s there” for the major players in the vaginal microbiome.

What we clearly don’t understand yet are the interactions between the organism components of the microbiota. Castro et al. (18) examined the adherence ability of two strains of *G. vaginalis*, one isolated from healthy conditions and the other from BV, and a strain of *L. iners* and *L. crispatus* isolated from healthy women. The BV *G. vaginalis* isolate had a higher capacity for binding cervical epithelial cells and could displace *L. crispatus* but not *L. iners*, whereas the healthy isolate had no effect on displacement. Furthermore, *L. iners* enhanced the adhesion of the BV strain but did not affect the healthy isolate. This study suggests some mechanisms for co-occurrence and exclusion between organisms in the vagina. The cumulative effect of co-occurrence and exclusion amongst all the organisms in an environment can be used to build predicted ecological networks describing the interactions of the community. Recent attempts have been made to use high-throughput 16S sequence data to predict these relationships with interaction networks (19, 20).

In terms of clinically defining BV, the current “gold standard” Nugent and Amsel criteria do not appear to encompass our new molecular understanding of the vaginal microbiome. As the cost of sequencing continues to drop, future clinical practices may use these techniques to profile the vaginal biota for diagnosis. Understanding the function (described below) and interactions between the host and microbes will make treatment planning more efficient and effective. The lack of progress from the medical industry has been stark, and clearly better diagnostic and more targeted therapeutics are needed to lower the immense burden of BV which impacts the quality of life of so many women (21). In the very near future, high throughput sequencing methods will make it feasible to obtain a vaginal microbiome profile of a patient during a doctor's visit. Armed with this information, the doctor will then be faced with its interpretation. Research is needed to determine which profiles require which, if any, intervention. For example, if a female is at higher than normal risk of STI through unsafe sex practices, and she has for example a *Prevotella* and *Lachnospiraceae* dominated microbiome, and if research shows this species increase risk of viral infection, then a treatment targeting depletion of *Prevotella*
and *Lachnospiraceae* would be optimal. Likewise for pregnant women, it may be that co-association of *Atopobium* and *Gardnerella* increase risk of preterm labour, in which case they might be targeted. At the same time as depleting pathogens, any therapy needs to help restore a 'healthy' state, in most cases the recovery of indigenous lactobacilli.

### 6.3 The role of *Lactobacillus iners*

*Lactobacillus iners* remains an enigma in the context of the vaginal microbiome. This organism was long missed by the failure of culture techniques, but molecular methods have unequivocally defined its predominance in the vagina. During BV, we detected *L. iners* at 10% relative abundance or more in nearly all cases, while other *Lactobacillus* species were rare (Chapter 5 and Section 1.4.2). Even in menopause, *L. iners* remains a major component as well as during acute VVC (Chapter 1 and Chapter 5). Very little description of this organism existed until we sequenced and described the genome of *L. iners* AB-1.

The reduced genome in comparison to other lactobacilli, had evidence of gene loss and an exceptionally high number of predicted horizontally acquired genes. The data overall suggest *L. iners* has a long evolutionary history of adaptation to the human vagina. An arsenal of unique adhesins were predicted, including a number of cell-wall anchored proteins and two fibronectin/fibrinogen binding proteins. A follow-up study I conducted with another graduate student confirmed that *L. iners* bound immobilized human fibronectin *in vitro* and the adhesion was significantly reduced after protease treatment (22). Fibronectin concentrations range from 90 to 140 pg/mL during menses to 3 to 5 pg/mL at the end of the menstrual cycle (22). As *L. iners* reportedly increases in abundance during menses while *L. crispatus* is less abundant (23), the fibronectin-binding protein may represent an adaptation for *L. iners* to persist during fluctuations in the environment.

We were the first to predict a cholesterol-dependent cytolysin in the genome of *L. iners*, and later reported this gene was upregulated 6-fold in BV compared to healthy (Chapter 4). The cytolysin shares considerable protein sequence similarity with another cytolysin present in *G. vaginalis* and known to be a key virulence factor (24). Rampersaud *et al.*
reported *in vitro* production of the *L. iners* cytolysin from a number of clinical isolates. The size of the pore is predicted to allow movement of small molecules from the cell cytoplasm to the extracellular space and vice versa. One possibility of why the organism uses this compound is to release nutrients from the human cells. Some researchers have interpreted this to mean that *L. iners* lyses cells and therefore participates in the onset of BV (26), and more likely AV in which lysed epithelial cells are present (27).

Although lactobacilli reportedly don’t require iron for cellular processes like most organisms, the genome sequence of *L. iners* revealed a predicted ferrochelatase not present in related lactobacilli. These features are usually described as 'virulence' factors for pathogenic organisms, but although *L. iners* could have a role in the aetiology of BV there is no evidence that it is a truly pathogenic organism.

The transcriptome profile of *L. iners* presented in Chapter 4 is the first general comparison of genomic content for this species. As shown in Figure 4-2, approximately 60% of the mapped coding sequences were conserved across 13 genomes while the remaining 40% was variable. Several highly differential coding sequences (e.g. CRISPR-associated) were only detected in one of the conditions (2 of the 4 samples). This offers another explanation for *L. iners’* persistence, as perhaps there are different strains more adapted to BV while other strains preferentially colonize healthy conditions. However, we were not able to assemble the CRISPR locus in the genome of *L. iners* AB-1 and similar assembly problems in the other strains could mean the differences are due to mapping to the reference sequences. Future studies using transcriptome or metagenome sequencing of different vaginal conditions with *L. iners* would provide information on differential strain abundance. Of note, although the sample size in this study is small, the strength of the data was such that the findings could be carried forward to larger comparative groups. Indeed, I am currently completing an analysis of 24 transcriptomes and with Amy McMillan comparing these to metabolomic profiles, and the results confirm the 4-subject study, plus reveal associations between bacterial types and their by-products. This work will be completed within the next few months but was not sufficiently developed for inclusion in this thesis.
Clearly *L. iners* is a vaginal organism that warrants more scrutiny, and indeed 45 of the 75 papers reporting on the organism in PubMed have come out in the past four years, compared to 30 papers between its discovery in 1999 and 2008. The differential transcriptomics has provided clues for the changing environment from a healthy to BV microbiota and *L. iners* may have a complex symbiotic relationship with its host ranging from mutualistic to pathogenic. Furthermore, it may be that different clones of *L. iners* encode different host outcomes.

### 6.4 What are the bacteria doing?

We described the first transcriptional analysis of the vaginal microbiome in Chapter 4 using metaRNAseq. The power of this study is limited by the sample size of four samples (for logistical and cost reasons), but we specifically chose and developed methodologies that would find robust, conserved results and to avoid over fitting the data. To overcome the obvious difference in gene content between samples due to differing taxa membership (Figure 4-1) the community-level analysis was performed at the functional level and not on individual genes belonging to individual species. We accomplished this by grouping the mapped genes by sequence identity and then assigning a broader functional classification to the group of genes with SEED subsystem annotation. For assessing differential expression, we grouped all the reads belonging to the same SEED functional subsystem for comparison. This accounts for redundancy in gene functions that may be constant despite population changes. For example, Gajer *et al.* (11) in reporting the vaginal microbiota and metabolome using $^1$H NMR (nucleic magnetic resonance) spectroscopy over time showed that a shift from *L. iners* abundance to *L. crispatus* abundance had no effect on the total metabolome. In their case, metagenomic quantification of the gene abundance would likely show a large change in gene content during this population shift despite the apparent functional conservation. In our study, we noted several functionally conserved features that were differential despite discordance in the organism contributing to the function (Figure 4-1). This presents an interesting ecological hypothesis that several species have overlapping niches in the vagina and can fulfill the same functional role.
The statistical framework, developed specifically for this study but applicable to any high-throughput sequencing experiment (28), reports significantly different features that are consistently higher in one defined condition or another (healthy or BV). Therefore, although this provides a level of confidence for the reported data, in order to strengthen conclusions drawn from four samples, we have performed a follow-up to Chapter 4 that included an additional 24 vaginal samples in addition to the original four (Figure 6-1).
Figure 6-1. Summary of mRNA and V6 16S rRNA gene sequence data for 28 vaginal samples

The mRNA samples presented in Chapter 4 (marked with a filled circle) were compared to a larger dataset of 24 samples. All samples were re-mapped to an updated refseq database and an updated SEED subsys4 level database was used to assign function. **Heat map:** Expression of each functional group (SEED subsys4 level) as derived by ALDEEx is plotted. Euclidian distance was calculated between samples and the dendrogram represents complete linkage clustering. Nugent status (red=BV, grey=intermediate, blue=normal) is plotted below the heat map. **Bar plots:** The top bar plot is the fractional composition of the mapped mRNA reads for each sample. In contrast, the bottom bar plot is the fractional composition of V6 16S rRNA gene sequencing. 16S sequence data are not available for the samples presented in Chapter 4 (marked with a filled circle) and so the bars are hatched out.

Plotting the inferred subsys4 level expression of these data and performing hierarchical clustering suggested a major split between predominantly *Lactobacillus*-dominated samples and samples with mixed species. This corresponded with the Nugent BV status. Importantly, the four samples presented in Chapter 4 (marked with a closed circle in Figure 6-1) clustered with the additional 24 samples by BV status and microbiota composition. This indicates the similarity is being driven by the overlap in gene expression and is independent of the platform used for sequencing. Interestingly, clustering by expression data suggested two major clusters for Nugent BV/intermediate samples. One cluster (samples on the far left of Figure 6-1: 019A, 003A, 008B, 001A) had a relatively high abundance of lactobacilli reported by V6 sequencing (bottom bar plot) and included two of these samples with *L. gasseri/johnsonii* as the predominant species, while other BV samples had *L. iners, G. vaginalis, Prevotella, BVAB1*, and *Megasphaera*. Within the healthy/non-BV cluster, sub-clusters consisted of *L. iners* or *L. crispatus* dominated profiles. Although preliminary, these data support overlapping conserved functions within BV or healthy conditions. Within healthy, there are possible subclasses defined by the different dominating *Lactobacillus* species.

Many of the transcriptional changes associated with BV (reported in Chapter 4 and unpublished data from Figure 6-1) were related to central carbohydrate metabolism and production of different metabolic end products (lactic acid in healthy conditions, and succinate, butyrate, acetate, propionate in BV conditions). These products have been
reported in association with BV and shown to promote a pro-inflammatory response in vitro (29). Although up to 30-40% of women with a BV Nugent profile report no symptoms (3, 30), the production of these metabolic products by the bacteria may still activate an immune response resulting constitute an increased risk for infection.

6.5 Considerations for modulating the ecosystem

Antibiotic treatment for BV fails in approximately 30% of patients, and up to another 30% of patients with initial recovery have a recurrent episode within 3 months and over 50% within one year (31). In cases of recurrent BV and VCC the treatment and prevention strategies are woefully inadequate, yet no new therapies are on the horizon (32). The concept of replenishing the vaginal lactobacilli through probiotics has been discussed and developed since 1988 to prevent recurrent UTI (33) and more recently to resolve recurrence of BV (34).

Our study reported in Chapter 5, is the first on three counts: it shows that probiotic therapy can modulate the vaginal microbiome, while others have reported it does not do so to the gut (35); it showed that the clinical recovery matched restoration of indigenous lactobacilli in BV patients; and it was the first report of the effects of VVC on the vaginal microbiome. Intriguingly, the host recovery did not require the two probiotic strains to universally reach or dominate the vagina. On the contrary, their oral administration induced a restoration of L. iners and L. crispatus for reasons that are only conjecture at present. Unlike L. iners, strains GR-1 and RC-14 do not adhere to fibronectin (22), which might reduce their ability to colonize the vagina since they reach the location in lower numbers compared to when they are intravaginally instilled (36). The ability to reduce pathogen ascension from the rectum to the vagina along the perineum might reduce pathogen seeding and thus allow better recovery of the indigenous lactobacilli (37). Another factor might be the well-known ability of lactobacilli to enhance mucosal immunity. Two previous studies from our laboratory showed that an aberrant vaginal microbiota may reduce the effectiveness of the immune response and the instillation of probiotic lactobacilli might to some extent restore the host's ability to counter pathogens (38, 39). There are several hypotheses for the role of the immune system in BV. Vaginal epithelial cells produce antimicrobial compound (defensins, lectins, proteases) and use
Toll-like receptors to recognize the microbial components in the environment and signal the immune system (40). The absence of an inflammation response in BV is considered an enigma (41), but is likely due to suppression of the immune response by the bacteria (42). The alterations in innate immunity has been proposed as a factor for the development of adverse effects, like pre-term labour, in some women with BV and not others (42). The cross-talk between host and microbes could have enormous effects on the development and resolution of conditions like BV and VVC, and changes in the microbiota.

The ability to enhance the host's defenses while also depleting pathogens, represent a different approach to those taken for the past 50 years where administration of antimicrobial agents have been designed solely to kill microbes, disregarding the implications for restoring homeostasis. The fact that in some cases probiotic therapy alone can better cure BV than antibiotics (43) further emphasizes that such approaches warrant more serious consideration. Of course, the long-term health of the female is equally as important as addressing an immediate acute infection. Understanding why some women never suffer from UTI, BV or VVC is an important future goal, and whether it is due to microbiome stability, immune efficiency and/or an ability to self-cure infection or another factor (44) will lead to improvement in care and quality of life. The idea that there are passenger organisms that simply adapt to fluctuations, and others that drive the change, is worth pursuing and is now feasible given the tools described herein.

6.6 Future directions

This thesis has put into place a number of techniques and experimental protocols that have provided great insight into the vaginal microbiome and Lactobacillus in particular. These now allow a series of investigations to be undertaken, which previously would not have been possible. The following outlines some areas that I think could prove fruitful.

High-throughput sequencing of the vaginal microbiome has revealed an unprecedented amount of information about the diversity, interactions, and functions of the bacterial residents. It is clear that BV is not one disease, and the possibility that differential functions of the microbes constitute different outcomes for the host needs to be further
investigated. Centres with high patient loads will be well placed to analyze BV patients and acquire sufficient numbers who fit into the different 'healthy' and 'BV' categories. Using transcriptomics merged with metabolomics, it will then be possible to look at what the bacteria are doing in real time, and by combining this with human microarrays and immune studies, to find out the effects on the host. This might possibly reveal early changes that increase the risk of preterm labour or viral infection. For example, studies using GCMS at Imperial College in London UK have shown that markers of cancer can be detected at sites a few centimeters from an actual tumour and that the identity of the tumour can be made along with its primary and/or secondary site (Dr. Jeremy Nicholson, personal communication). This was made possible by collecting thousands of metabolic readouts. The same should be possible for the vagina through collaborations with Dr. Ravel (St. Louis), Dr. Bocking (Toronto), Dr. Money (Vancouver) and others to collect and match profiles with disease outcomes and predict which ones will be problematic. In our transcriptomic and matching metabolomic studies, we have identified markers for BV such as odor compounds that likely are responsible for patients seeking medical attention. In other cases of BV if odor production is less, we will be able to determine whether risk of complications is also lowered or not. Such information is vital for caregivers to determine if patients require treatment or if therapy has been effective.

Longitudinal studies, like the one performed by Ravel's group, will help identify the normal day-to-day fluctuations of the microbiota in response to exogenous disturbances, and differentiate benign changes with ones that have pathological implications.

Other than the bacteria, the vaginal microbiome consists of the viruses and eukaryotic organisms either transiently present during infection (Trichomonas, Candida) or established members of the community. We reported the bacterial microbiota during VVC was undisturbed compared to normal, but the coexistence of Candida causing host response and symptoms may still affect the response of the bacteria in the environment, and therefore affect the host. Transcriptional profiling of L. iners revealed the upregulation of several phage defense systems, notably CRISPR regions, that suggest viral influence on the population. Indeed, the idea has been proposed that Lactobacillus-targeting phage could be responsible for their depletion leading to BV (45). Another
hypothesis is that bacteriophage adherent to the human mucosa can act as a non-human component of immunity (46). Viral diversity is only beginning to be explored with metagenomic techniques in the human gut (47), and these approaches could be extended to the vaginal microbiome. Similarly, the eukaryotic microbiota of the vagina is not well defined with only one study to date (48) using high-throughput sequencing to examine the diversity of the vaginal mycobiome.

Current two-species biofilm models have already shown the effects of exogenously applied lactobacilli (49), but growing a multispecies biofilm or culture is limited with standard laboratory techniques. An in vitro artificial model, like the “robogut” used to model the gut microbiota (50) could provide a useful tool to develop and test concepts. Currently, apart from primates that are expensive to work with, there is no animal model to test the effects of exogenous disturbances on different vaginal communities. Many scientists use mouse models, but these are often too far removed from the human system, and therefore conclusions are difficult to transfer.

The challenge remains to turn the hypothesis-generating results from microbiome studies into predictive values for health and disease. In the case of the vaginal microbiome, we need to be able to correlate the bacterial actions to outcomes like pre-term birth, development of BV, and acquisition of infections diseases. The data presented in this thesis begin to answer some of these biological questions, and the tools developed for high-throughput sequence analysis will further our understanding in future studies. The reliance on pharmaceutical and diagnostic companies coming up with new products has so far proved hollow, and meanwhile millions of women around the world continue to receive sub-optimal care. The use of probiotics, especially through food, could be a faster route to patient care as it does not require extensive drug studies and animal toxicity assessment. Our studies clearly show this has benefits for restoring and maintaining vaginal health even when administered with antimicrobial agents. Indeed, when administered to HIV patients in the Tanzanian community which I studied, there have been gut and vaginal benefits, including reducing recurrence of BV (51-53). Our findings indicate that the vaginal microbiome of HIV patients is not significantly different from those of other women in terms of the types of bacteria present, and others have shown
that they too can benefit from probiotics (54). Whether or not such probiotics should include *L. iners* remains to be determined, and with the species being difficult to propagate, the manufacture of such products may prove difficult. But, more studies will needed to investigate clonal differences between strains and see if some are better predictors of health than others.

Much of my thesis was involved in developing computational tools for these studies. This is far from complete, and improving the quality of samples, the depth of coverage, the degree to which silent or unidentified genes, or transcriptional regulators like riboswitches and anti-sense transcripts are involved in lactobacilli colonization and host defense is worthy of further effort (55).

At the end of the day, the driving force for these studies is the health of women. Given the impact on quality of life, reproductive health and outcomes of pregnancy caused by the vaginal microbiota, it is fundamentally essential that we continue to apply the findings of this thesis to improving female well-being. I feel confident that this thesis has at least provided some pathways to making that a reality.

### 6.7 References

1. van Leeuwenhoek A (1800) *The Select Works of Anthony Van Leeuwenhoek: Containing His Microscopical Discoveries in Many of the Works of Nature*, (translator,


Appendices
Appendix A: Copyright from PLoS ONE

Open-Access License

No Permission Required

PLOS applies the Creative Commons Attribution License (CCAL) to all works we publish (read the human-readable summary or the full license legal code). Under the CCAL, authors retain ownership of the copyright for their article, but authors allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in PLOS journals, so long as the original authors and source are cited. **No permission is required from the authors or the publishers.**

In most cases, appropriate attribution can be provided by simply citing the original article (e.g., Kaltenbach LS et al. (2007) Huntington Interacting Proteins Are Genetic Modifiers of Neurodegeneration. PLOS Genet 3(6): e82. doi:10.1371/journal.pgen.0030082). If the item you plan to reuse is not part of a published article (e.g., a featured issue image), then please indicate the originator of the work, and the volume, issue, and date of the journal in which the item appeared. For any reuse or redistribution of a work, you must also make clear the license terms under which the work was published.

This broad license was developed to facilitate open access to, and free use of, original works of all types. Applying this standard license to your own work will ensure your right to make your work freely and openly available. Learn more about open access. For queries about the license, please contact us.
Appendix B: Copyright from PNAS

Author Rights and Permissions
Frequently Asked Questions

1. Since I retain copyright but give PNAS an exclusive license to publish my article, what rights do I have?

As a PNAS author, you and your employing institution or company retain extensive rights for use of your materials and intellectual property. You retain these rights and permissions without having to obtain explicit permission from PNAS, provided that you cite the original source:

- The right to post a PDF of your article on your Web site or that of your employer's institution (provided that the institution is nonprofit).
- The right to make electronic or hard copies of articles for your personal use, including classroom use, or for the personal use of colleagues, provided those copies are not for sale and are not distributed in a systematic way outside of your employing institution.
- The right to post and update a preprint version of your article on a public electronic server such as the Web. See the information on electronic preprints below.
- The right to permit others to use your original figures or tables published in PNAS for noncommercial and educational use (i.e., in a review article, in a book that is not for sale), provided that the original source is cited. Third parties need not ask PNAS for permission to use figures and tables for such use.
- The right, after publication in PNAS, to use all or part of your article in a printed compilation of your own works, such as collected writings or lecture notes.
- If your article is a "work for hire" made within the scope of your employment, your employer may use all or part of the information in your article for intracompany use.

- The right to include your article in your thesis or dissertation.
- The right to present all or part of your paper at a meeting or conference, including ones that are webcast, and to give copies of your paper to meeting attendees before or after publication in PNAS. For interactions with the media prior to publication, see the PNAS Policy on Media Coverage.
- The right to publish a new or extended version of your paper provided that it is sufficiently different to be considered a new work.
- The right to expand your article into book-length form for publication.
- The right to reuse your original figures and tables in your future works.
- Patent and trademark rights or rights to any process or procedure described in your article.
Appendix C: Copyright from Microbiome/BioMed Central

BioMed Central Open Access Charter

Every peer-reviewed research article appearing in any journal published by BioMed Central is 'open access', meaning that:

1. The article is universally and freely accessible via the Internet, in an easily readable format and deposited immediately upon publication, without embargo, in an agreed format - current preference is XML with a declared DTD - in at least one widely and internationally recognized open access repository (such as PubMed Central);

2. The author(s) or copyright owner(s) irrevocably grant(s) to any third party, in advance and in perpetuity, the right to use, reproduce or disseminate the research article in its entirety or in part, in any format or medium, provided that no substantive errors are introduced in the process, proper attribution of authorship and correct citation details are given, and that the bibliographic details are not changed. If the article is reproduced or disseminated in part, this must be clearly and unequivocally indicated.

BioMed Central is committed permanently to maintaining this open access publishing policy, retrospectively and prospectively, in all eventualities, including any future changes in ownership.
Appendix D: Ethics approval for study presented in Chapter 2

Office of Research Ethics
The University of Western Ontario

Use of Human Subjects - Ethics Approval Notice

Principal Investigator: Dr. G. Reid
Review Number: 16182E
Review Date: February 09, 2010
Review Number: 2
Revision Number: 2

Protocol Title: The role of vaginal microbiota in symptoms of vaginal dryness in postmenopausal women
Department and Institution: Microbiology & Immunology, St. Joseph’s Health Care London
Sponsor: Kimberly Clark Corporation

Ethics Approval Date: February 09, 2010
Expiry Date: December 31, 2010

Documents Reviewed and Approved: Revised participant recruitment and number of study participants, Letter of Information and consent (version 1.5, Feb 2, 2010), Pamplet (version 3, Feb 2, 2010), Poster (version 3.0), Poster (version 3.1), Study Ad (Version 1.0), Study Ad (Version 1.1)

Documents Received for Information:

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB’s as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB’s periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the HSREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of monitor, telephone number). Expedited review of minor change(s) in ongoing studies will be considered. Subjects must receive a copy of the signed information/consent documentation.

Investigators must promptly also report to the HSREB:
   a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
   b) all adverse and unexpected experiences or events that are both serious and unexpected;
   c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to this office for approval.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

Chair of HSREB: Dr. Joseph Gilbert
FDA Ref. #: IRB 00000940
Appendix E: Ethics approval for study presented in Chapter 3

Office of Research Ethics
The University of Western Ontario

Use of Human Subjects - Ethics Approval Notice

Principal Investigator: Dr. G. Reid
Review Number: 16709E
Review Date: December 16, 2009
Protocol Title: Investigating the adhesion and gene expression of Lactobacillus iners in the vaginal environment.
Department and Institution: Microbiology & Immunology, St. Joseph's Health Care London
Sponsor:
Ethics Approval Date: December 22, 2009
Expiry Date: May 31, 2010

This is to notify you that the University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines; and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced study on the approval date noted above. The membership of this REB also complies with the membership requirements for REBs as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB’s periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the HSREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of monitor, telephone number). Expedited review of minor change(s) in ongoing studies will be considered. Subjects must receive a copy of the signed information/consent documentation.

Investigators must promptly also report to the HSREB:
- a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;
- b) all adverse and unexpected experiences or events that are both serious and unexpected;
- c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to this office for approval.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

Chair of HSREB: Dr. Joseph Gilbert
FDA Ref. #: IRB 00000340

UWO HSREB Ethics Approval - Initial
v.2008-07-01 (pdfApproval/notice/HSREB_initial) 16709E
Appendix F: Ethics approval for study presented in Chapter 4

Office of Research Ethics
The University of Western Ontario

Use of Human Subjects - Ethics Approval Notice

Principal Investigator: Dr. G. Reid
Review Number: 16183E
Review Date: July 16, 2010
Revision Number: 4

Review Level: Expedited

Protocol Title: The role of vaginal bacteria in toxic shock syndrome

Department and Institution: Microbiology & Immunology, St. Joseph’s Health Care London

Sponsor: Kimberly Clark Corporation

Ethics Approval Date: September 30, 2010
Expiry Date: June 30, 2010

Documents Reviewed and Approved: Revised study methods and participant recruitment. Recruitment poster.

Documents Received for Information:

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines, and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB's as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB's periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the UWO Updated Approval Request Form.

During the course of the research, no deviations from, or changes to, the protocol or consent form may be initiated without prior written approval from the HSREB except when necessary to eliminate immediate hazards to the subject or when the change(s) involve only logistical or administrative aspects of the study (e.g. change of monitor, telephone number). Expedited review of minor change(s) in ongoing studies will be considered. Subjects must receive a copy of the signed information/consent documentation.

Investigators must promptly also report to the HSREB:

a) changes increasing the risk to the participant(s) and/or affecting significantly the conduct of the study;

b) all adverse and unexpected experiences or events that are both serious and unexpected;

c) new information that may adversely affect the safety of the subjects or the conduct of the study.

If these changes/adverse events require a change to the information/consent documentation, and/or recruitment advertisement, the newly revised information/consent documentation, and/or advertisement, must be submitted to this office for approval.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

Chair of HSREB: Dr. Joseph Gilbert

FDA Ref: # IRB 000000940

This is an official document. Please retain the original in your files.
Appendix G: Ethics approval for study presented in Chapter 5

Comitê de Ética em Pesquisa do Centro de Saúde Escola da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo

OF. N° 224/05/COORD.CEP/CSE-FMRP-USP. 20/4/2005.

Senhora Professora,

Vimos na presente oportunidade comunicar que o Comitê de Ética em Pesquisa do Centro de Saúde Escola da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, em sua 38ª reunião ordinária, realizada em 19 de abril de 2005, analisou e aprovou o parecer do Senhor Relator, referente ao Projeto de Pesquisa “Efeito da utilização de culturas lácticas probióticas na microbiota vaginal de pacientes acometidas por infecções bacterianas e fúngicas”, protocolo 0146/CEP/CSE-FMRP-USP, que será conduzido por V. Sª e pelo orientando Rafael Chacon Ruiz Martinez, foi aprovado. Portanto, solicitamos que as alterações, referentes às solicitações feitas, sejam incorporadas a uma nova versão e encaminhada a este CEP.

Lembramos que em atendimento à Resolução 196/96, deverá ser encaminhado a este CEP o relatório final da pesquisa e a publicação de seus resultados.

Sendo o que tínhamos para o momento, despedimo-nos.

Atenciosamente,

Prof. Dr. Laercio Joel Franco
Coordenador do CEP/CSE-FMRP-USP

Ilma. Sra.
Profª Dra. Elaine Cristina Pereira de Martinis
Faculdade de Ciências Farmacêuticas de Ribeirão Preto-USP.
Appendix H: Ethics approval for study presented in Chapter 6

Use of Human Participants - Ethics Approval Notice

Principal Investigator: Dr. Gregor Reid
File Number: 100751
Review Level: Delegated
Approved Local Adult Participants: 100
Approved Local Minor Participants: 0
Protocol Title: Investigating small molecules produced by naturally occurring bacteria of the vaginal tract. - 18203E
Department & Institution: Schulich School of Medicine and Dentistry Microbiology & Immunology, Western University
Sponsor:
Ethics Approval Date: June 15, 2012 Expiry Date: June 30, 2016
Documents Reviewed & Approved & Documents Received for Information:

<table>
<thead>
<tr>
<th>Document Name</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revised Western University Protocol</td>
<td>Revised study methods, participant recruitment, eligibility of participants and administrative changes.</td>
</tr>
<tr>
<td>Other</td>
<td>Instrument - Participant Questionnaire</td>
</tr>
<tr>
<td>Other</td>
<td>Instrument - Vaginal Examination - Clinician Questionnaire</td>
</tr>
<tr>
<td>Revised Letter of Information &amp; Consent</td>
<td>2012/04/26</td>
</tr>
</tbody>
</table>

This is to notify you that The University of Western Ontario Research Ethics Board for Health Sciences Research Involving Human Subjects (HSREB) which is organized and operates according to the Tri-Council Policy Statement: Ethical Conduct of Research Involving Humans and the Health Canada/ICH Good Clinical Practice Practices: Consolidated Guidelines, and the applicable laws and regulations of Ontario has reviewed and granted approval to the above referenced revision(s) or amendment(s) on the approval date noted above. The membership of this REB also complies with the membership requirements for REB’s as defined in Division 5 of the Food and Drug Regulations.

The ethics approval for this study shall remain valid until the expiry date noted above assuming timely and acceptable responses to the HSREB’s periodic requests for surveillance and monitoring information. If you require an updated approval notice prior to that time you must request it using the University of Western Ontario Updated Approval Request Form.

Members of the HSREB who are named as investigators in research studies, or declare a conflict of interest, do not participate in discussion related to, nor vote on, such studies when they are presented to the HSREB.

The Chair of the HSREB is Dr. Joseph Gilbert. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

This is an official document. Please retain the original in your files.
Curriculum Vitae

Name: Jean M. E. Macklaim

Education:
The University of Western Ontario 2008-2013
London, Ontario, Canada
Ph.D Biochemistry

University of Western Ontario 2004-2008
London, Ontario, Canada
Hons. BSc Biology, Specialization in Genetics

Awards and honours:

Ontario Graduate Scholarship (OGS) 2011-2013
Queen Elizabeth Scholarship II (declined) 2012
Ontario Graduate Scholarship in Science and Technology (OGSST) 2010-2011

1st place poster presentation, Infection and Immunity Research Forum 2012
2nd place platform presentation, Lawson Research Day 2011

Publications:


**Manuscripts submitted or in preparation:**


